Software Estimation: Practical Insights & Orphean Research Issues

Alain Abran

École de Technologie Supérieure, University of Québec, Montréal, Canada

alain.abran@etsmtl.ca

9th International Conference Software Engineering Advances

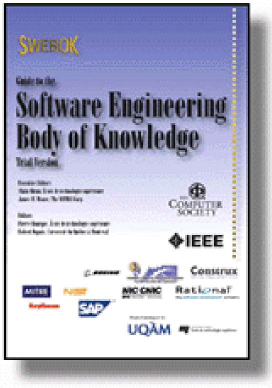
October 12-16, 2014 - Nice (France)

Alain Abran

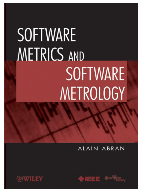
20 years

20 years

- **Development**
- **Maintenance**
- **Process Improvements**



+ 35 PhD



ISO:
19761, 9216,
25000, 15939,
14143, 19759

List of topics

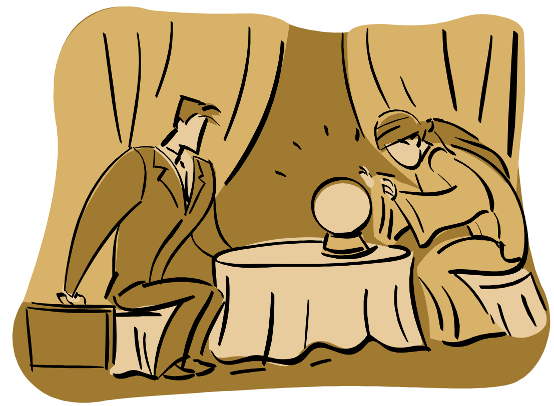
1. Estimation: Craft or Engineering?
2. The estimation phases
3. Economic concepts for estimation models
4. Orphean research issues

List of topics

1. **Estimation: Craft or Engineering?**
2. The estimation phases
3. Economic concepts for estimation models
4. Orphean research issues

(Software) Estimation

Or?



Estimation expectations

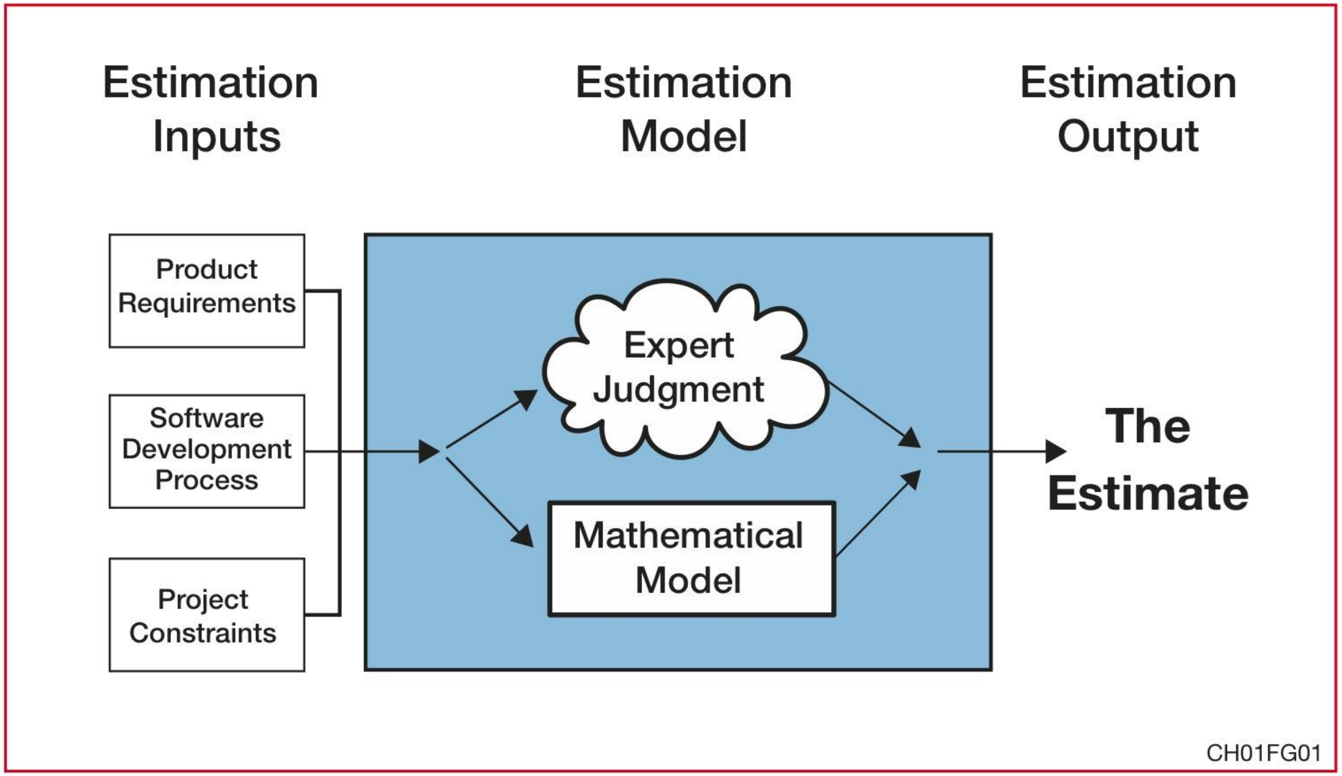


Figure 1.1 Common view of an estimation process.

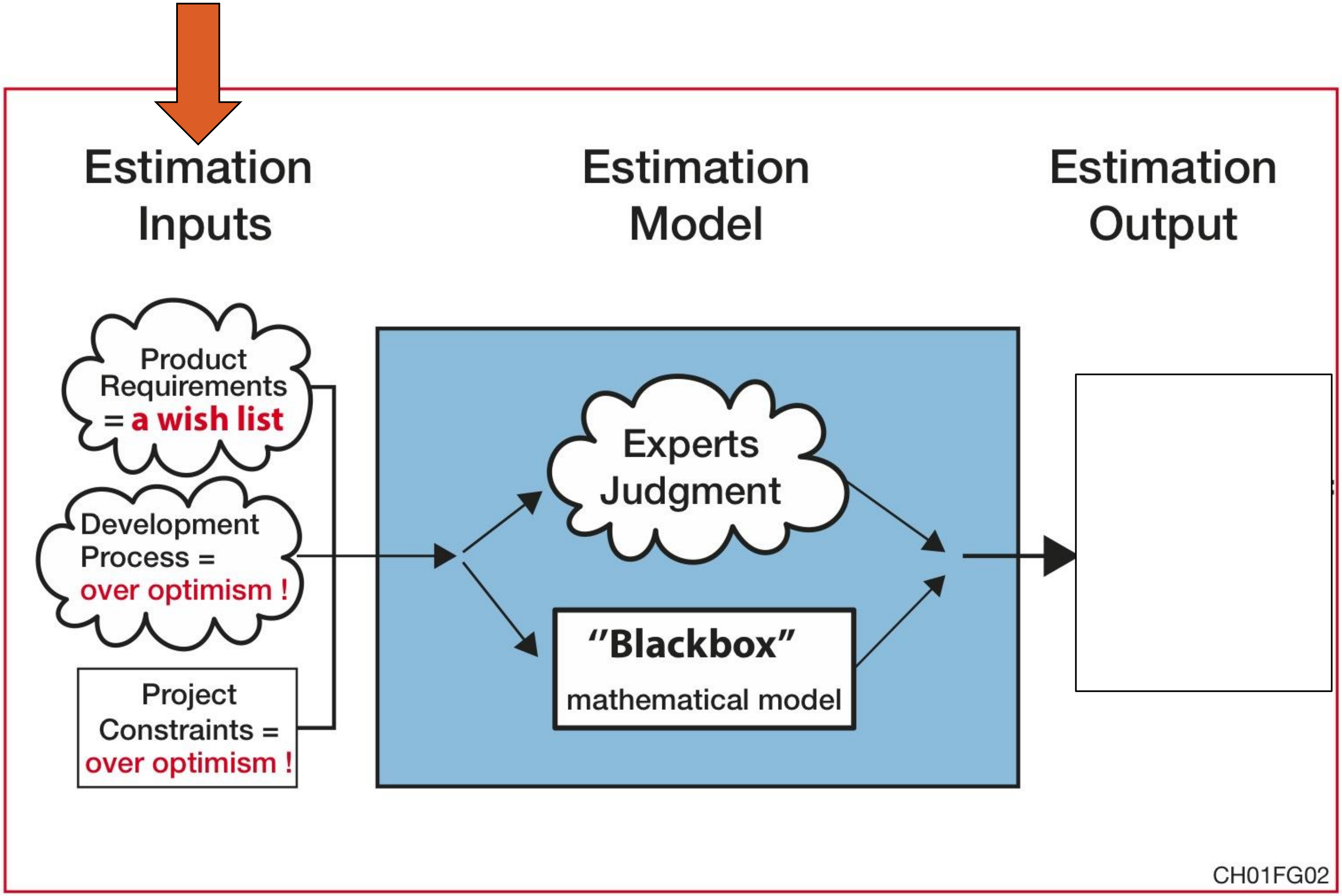


Figure 1.2 Some poor estimation practices observed in industry.

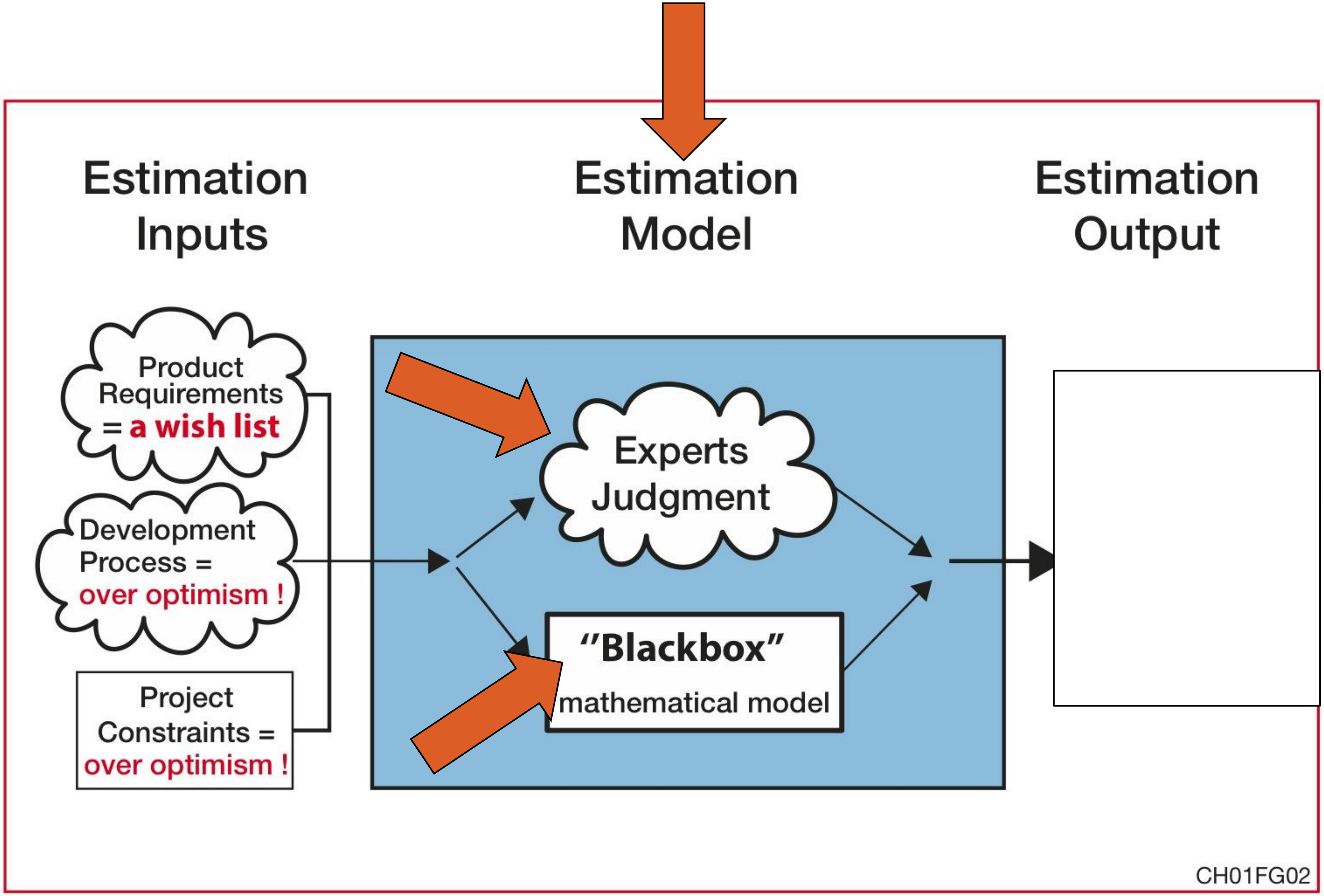


Figure 1.2 Some poor estimation practices observed in industry.

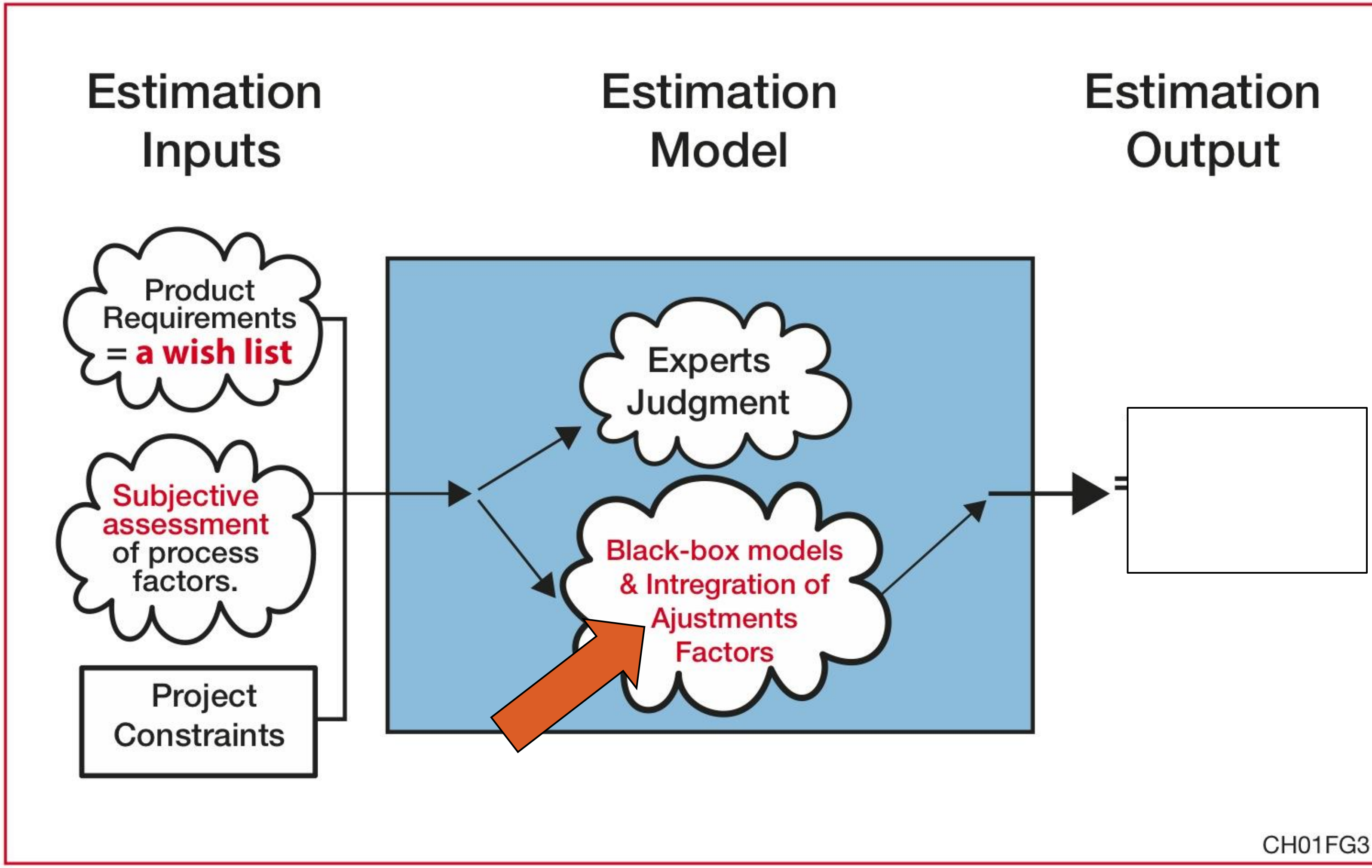


Figure 1.3 Some of the worst estimation practices.

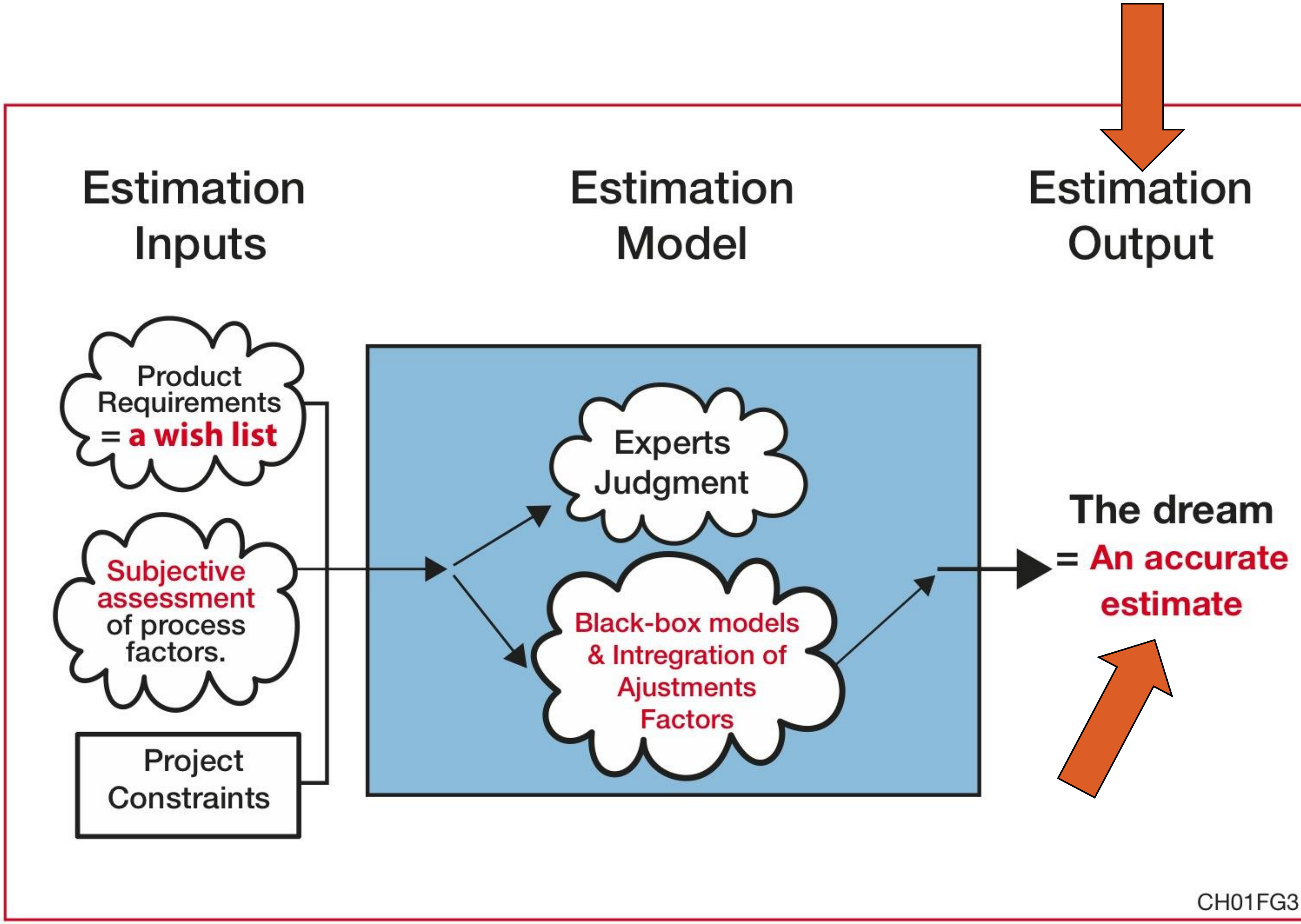
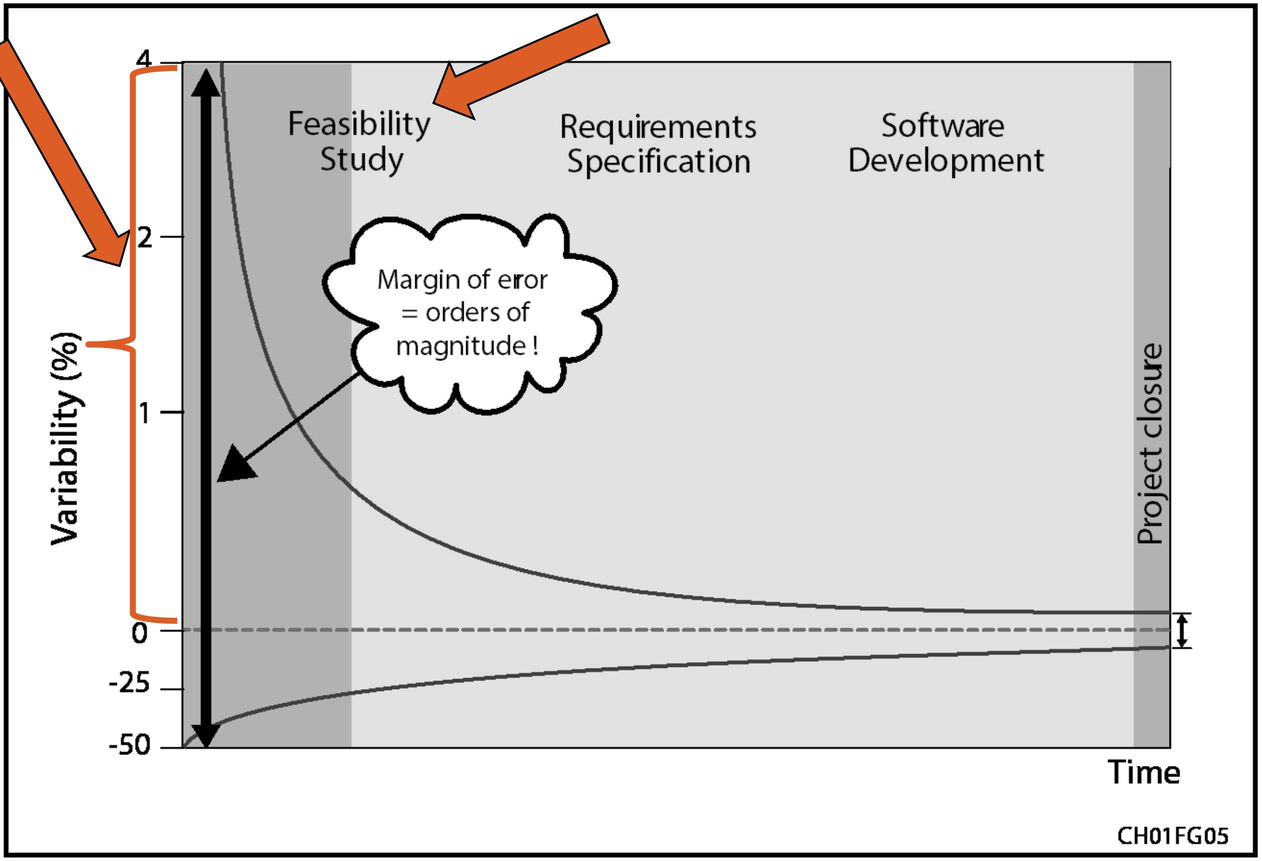


Figure 1.3 Some of the worst estimation practices.

Imprecise Inputs at Feasibility Analysis – Much Greater Error Range



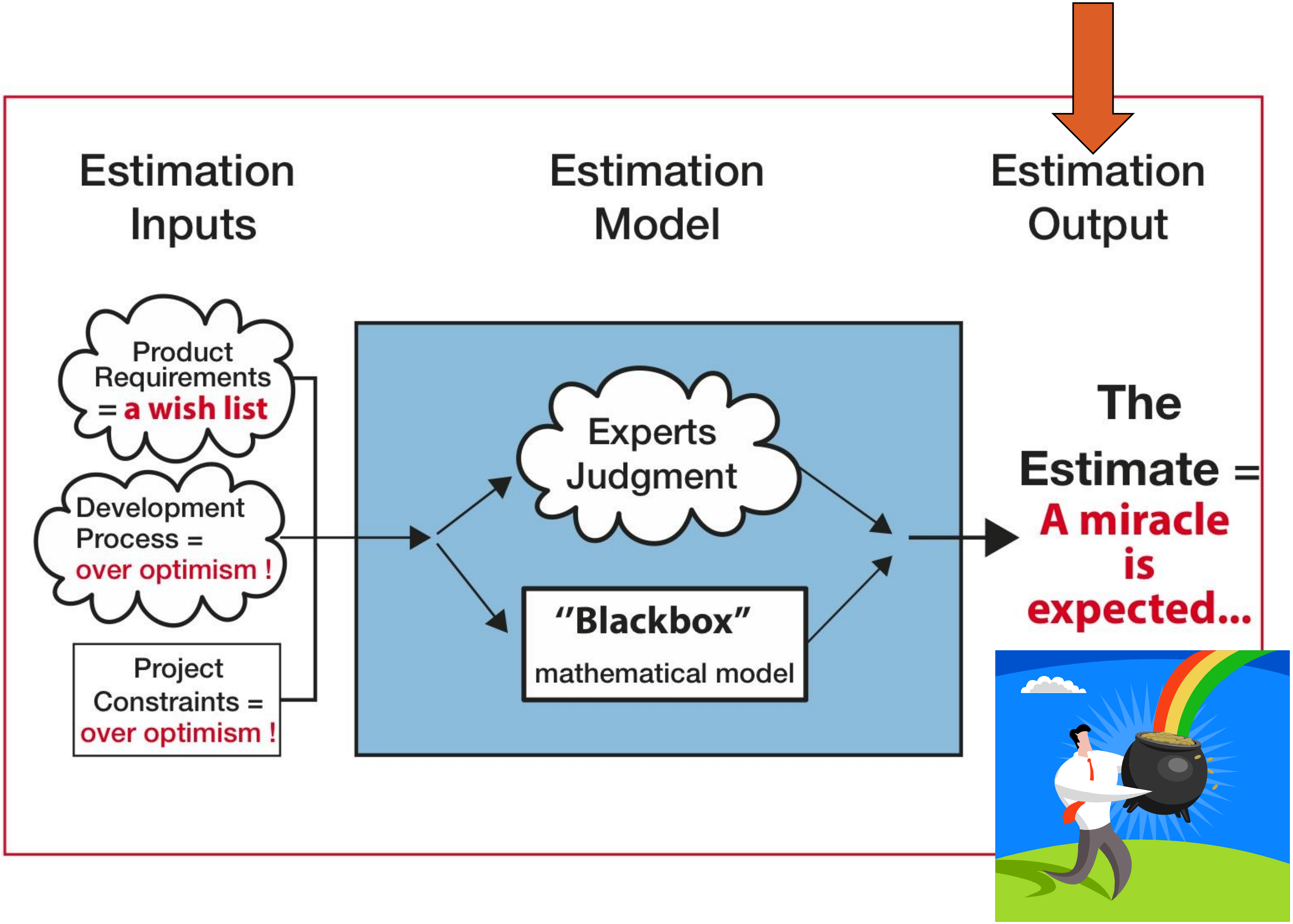


Figure 1.2 Some poor estimation practices observed in industry.

A look at the most-known estimation approach:

➡ The 'COCOMO-like' approach with its 'cost drivers' where:

Effort = F(Size,+15 'cost drivers')

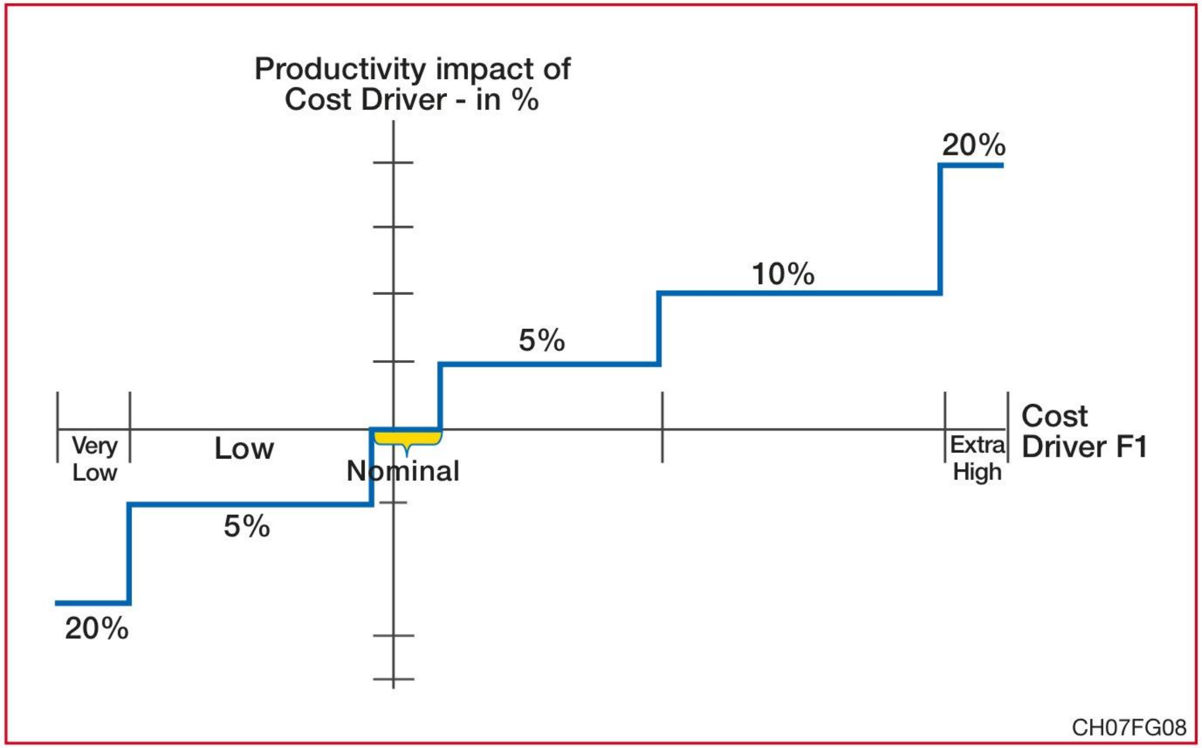
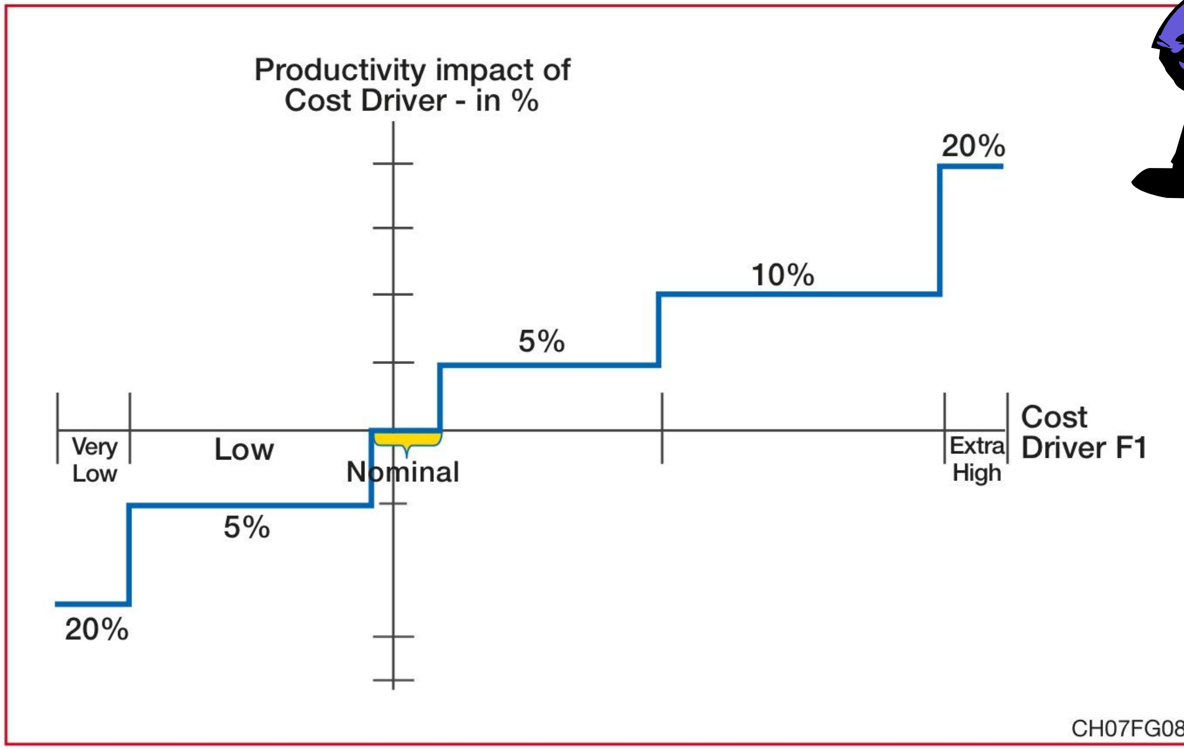


Figure 7.8 A step-function estimation model-
with irregular intervals.



Estimation by 'experts'

Figure 7.8 A step-function estimation model- with irregular intervals.

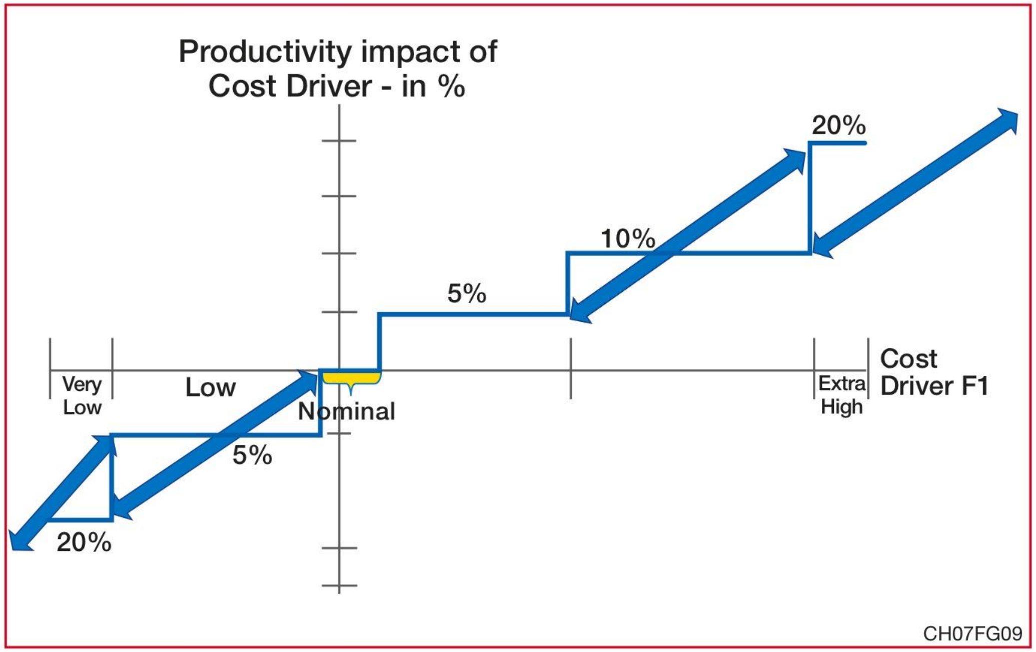


Figure 7.9 Approximation of step-funtion productivity models with irregular intervals.

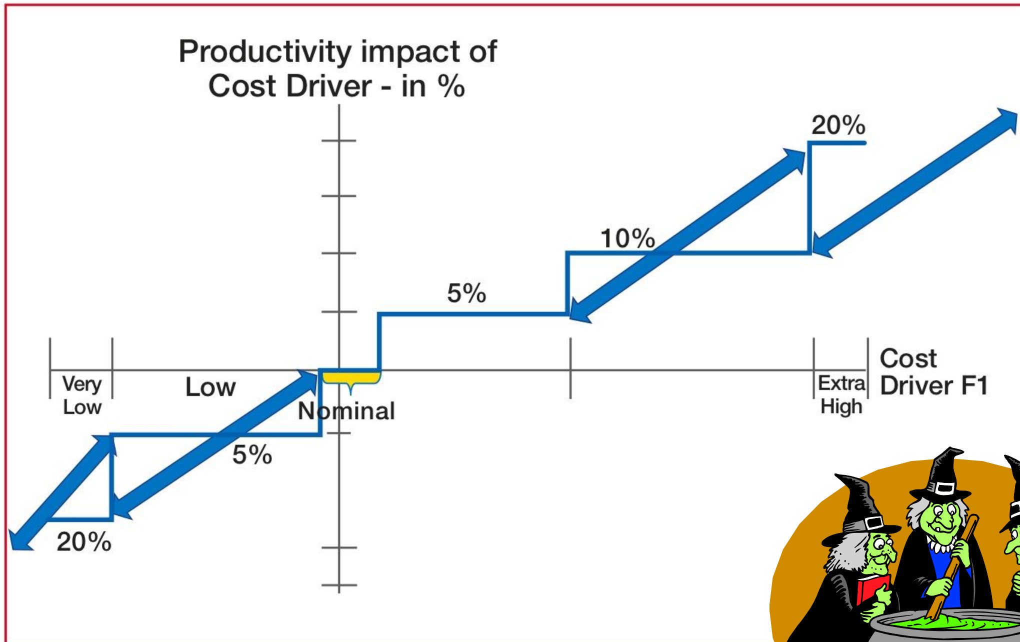
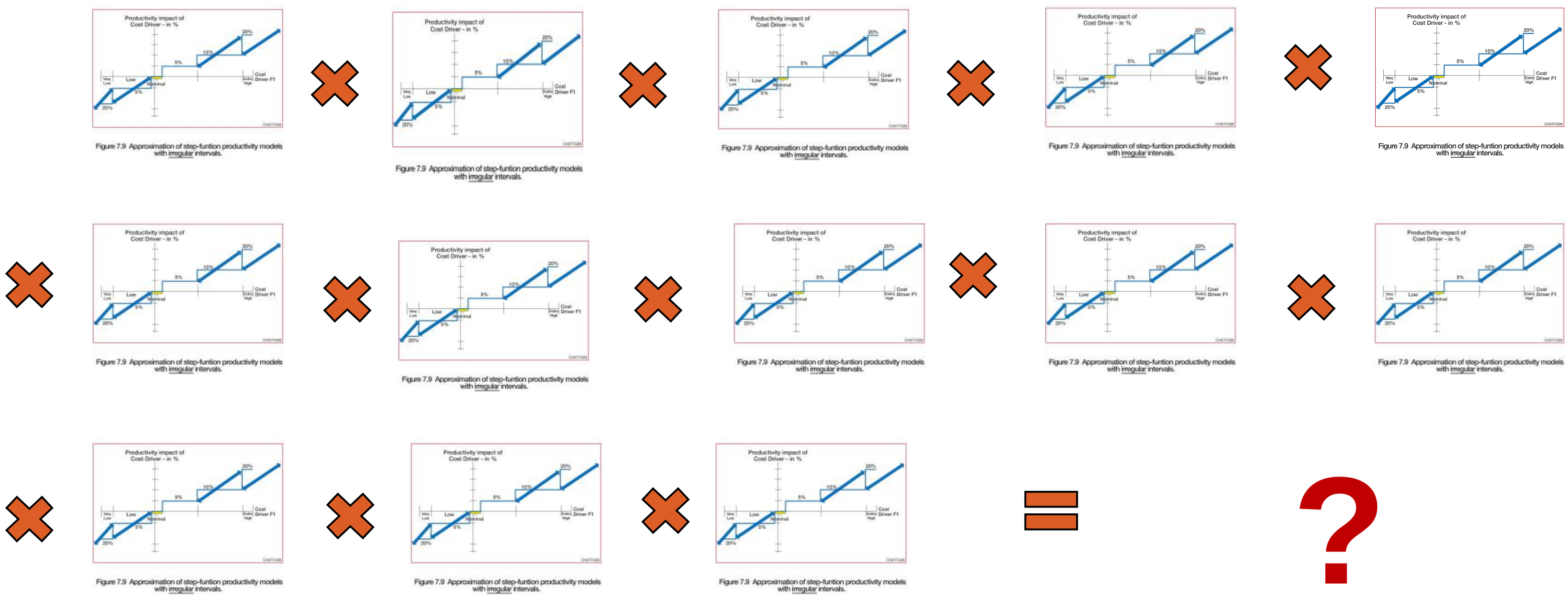


Figure 7.9 Approximation of step-funtion productiv
with irregular intervals.

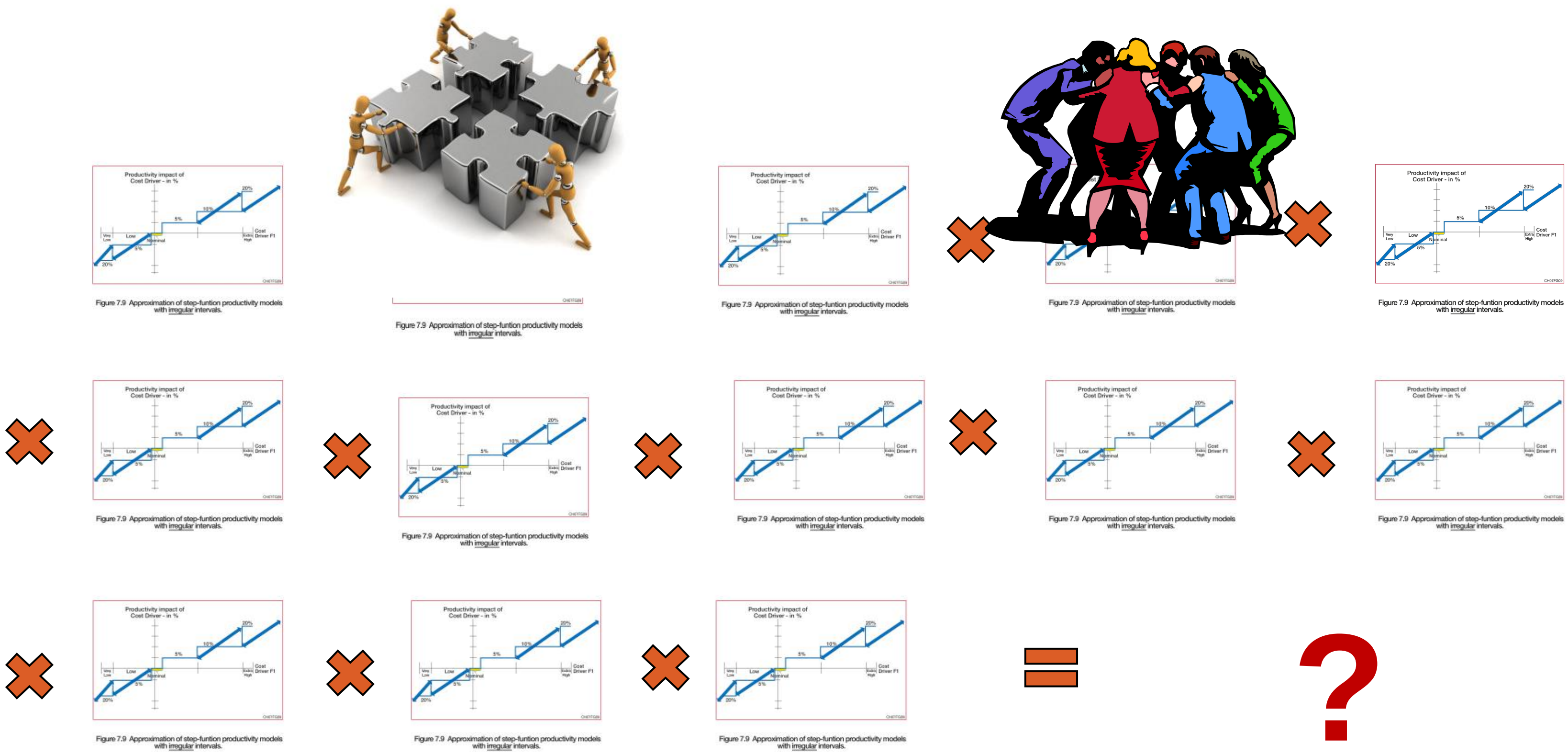
Each COCOMO cost driver =

an estimation sub-model with unkown quality & large errors

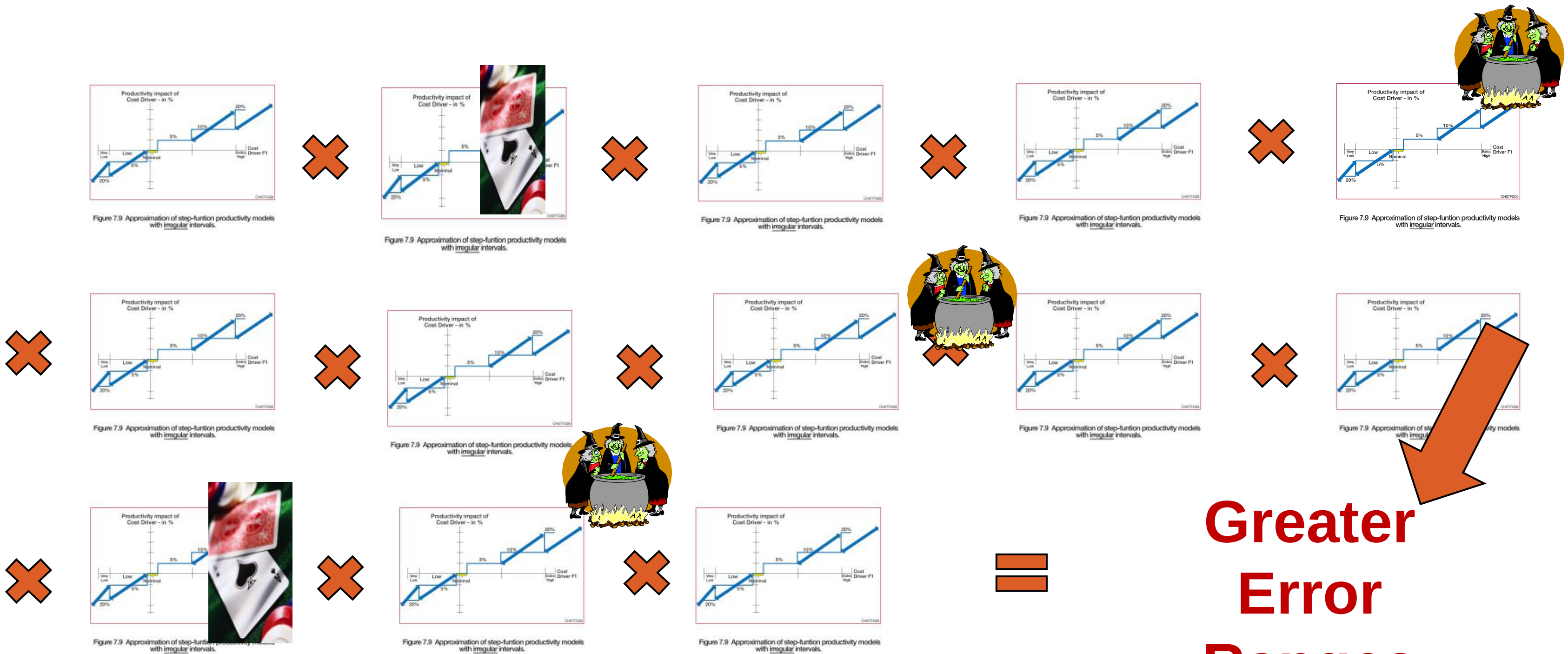
COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions)



COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions)

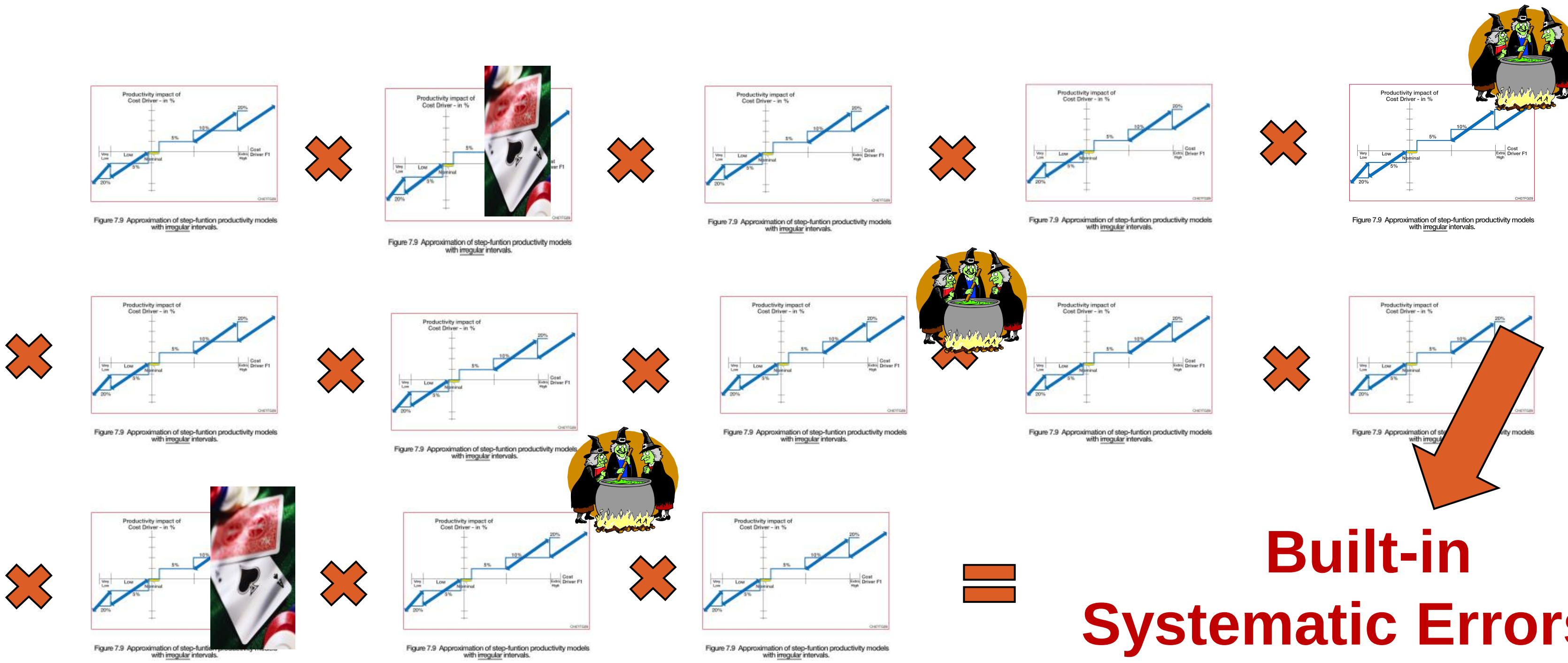


COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions)
of unknown quality combined into a single number!



Greater Error Ranges

COCOMO-like estimation models:
Effort is a function of (Size & +15 step-functions)
of unknown quality combined into a single number!



Built-in Systematic Errors & Error Propagation

Kemerer 1987 on COCOMO81

Small scale replication study - 17 projects

	Basic Exponential on Size	Intermediate & 15 cost drivers	Detailed & 4 project phases
R^2 (max=1.0)	**0.68**	**0.60**	**0.52**
MMRE **(mean magnitude of relative errors)**	**610%**	**583%**	**607%**

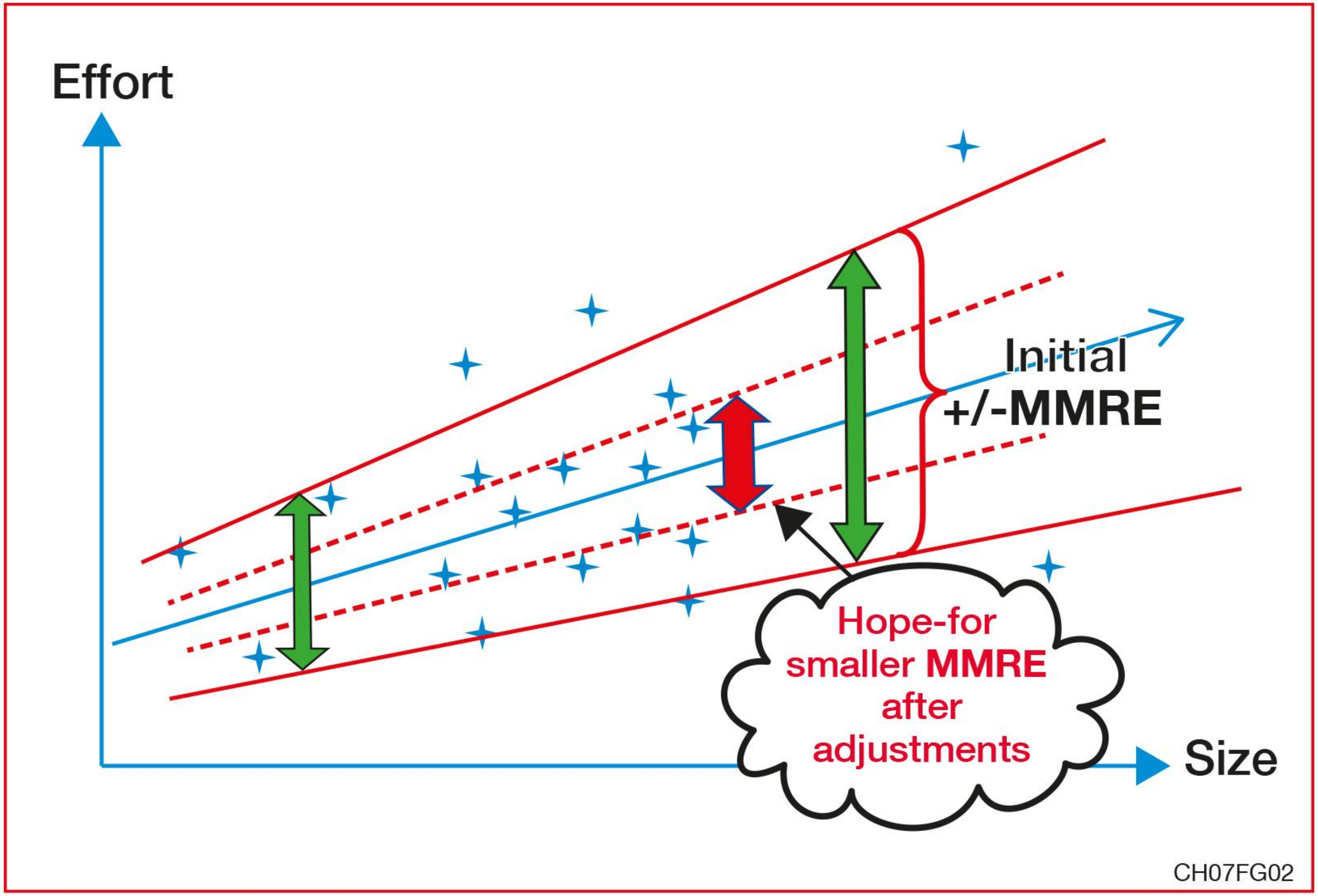


Figure 7.2 Desired impact.

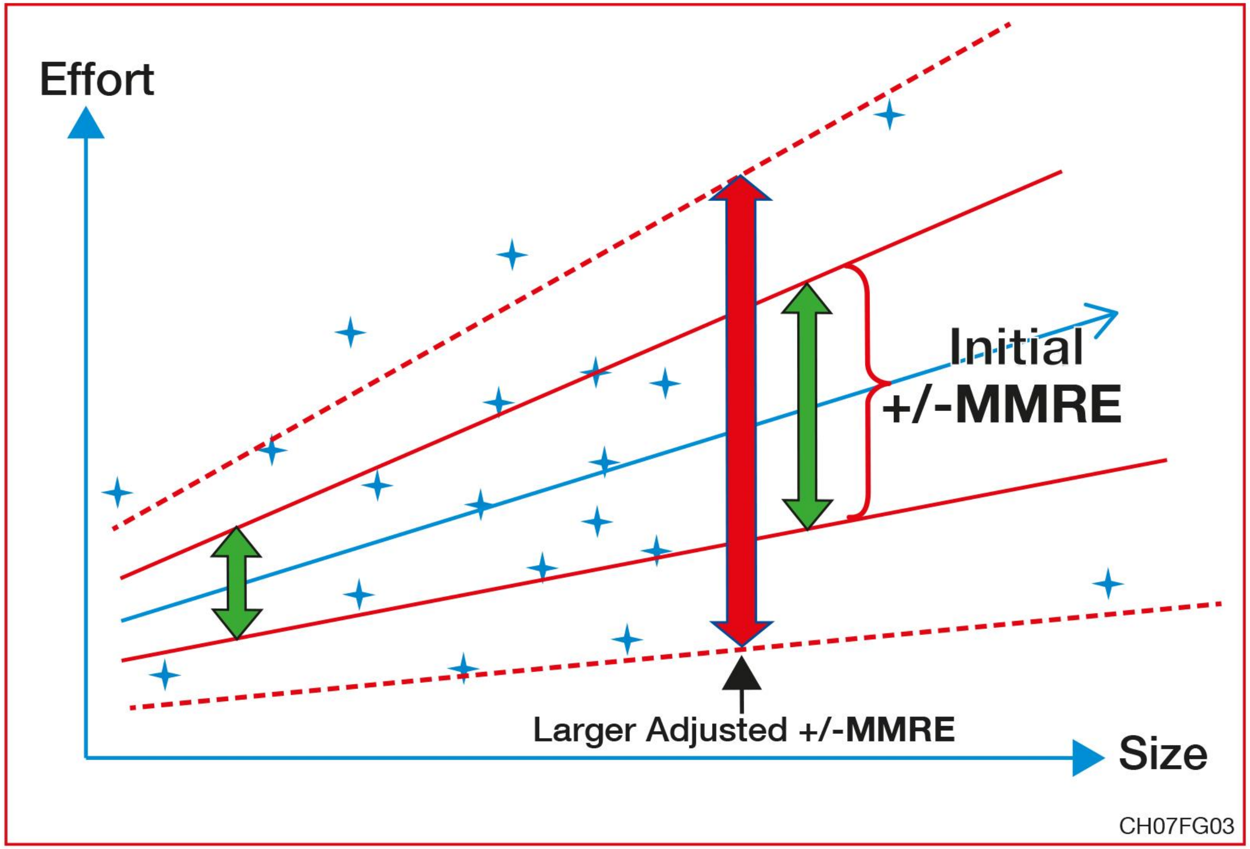


Figure 7.3 Plausible Greater Impact of Adjustments to Estimates.

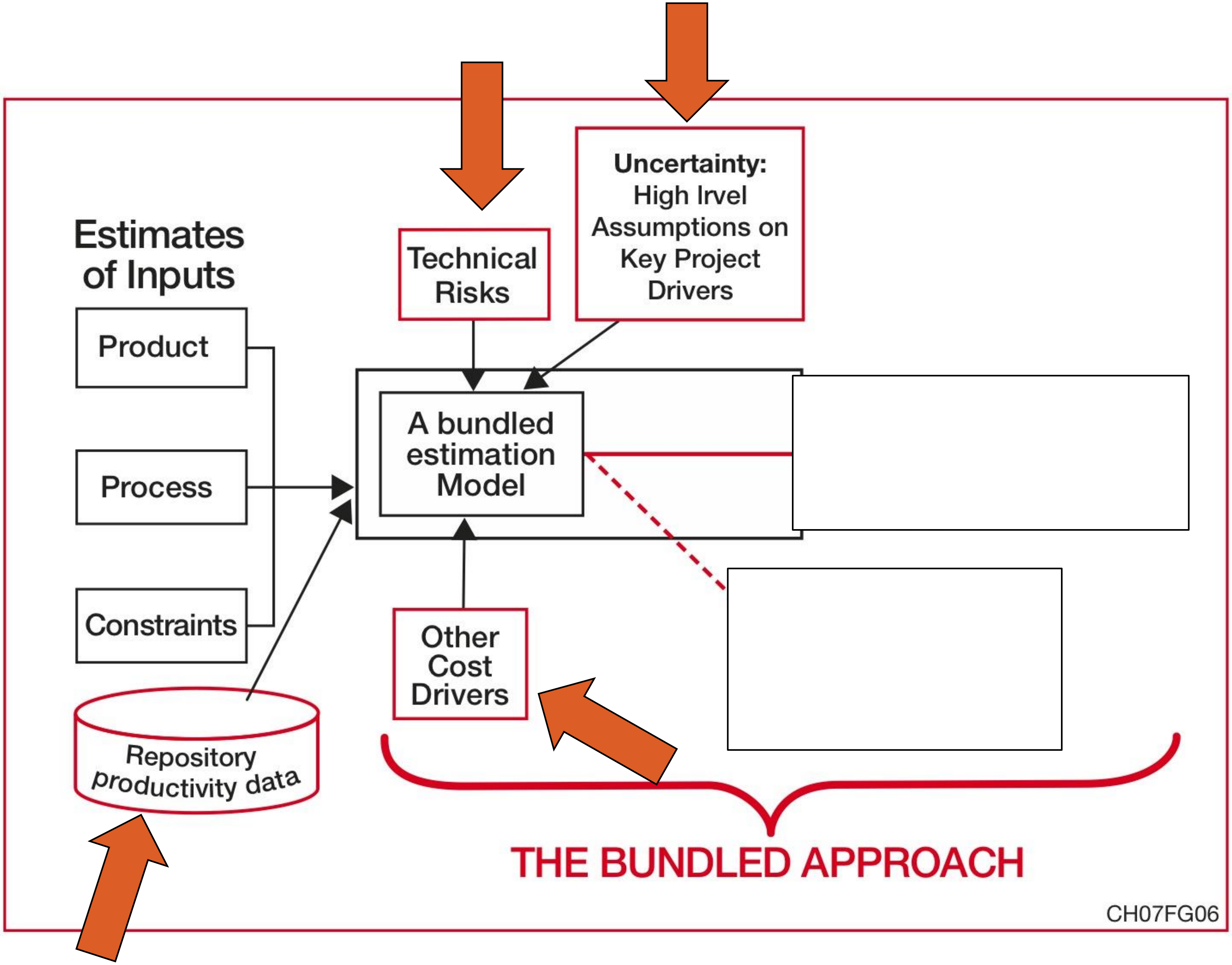


Figure 7.6 Estimation - The Bundle Approach.

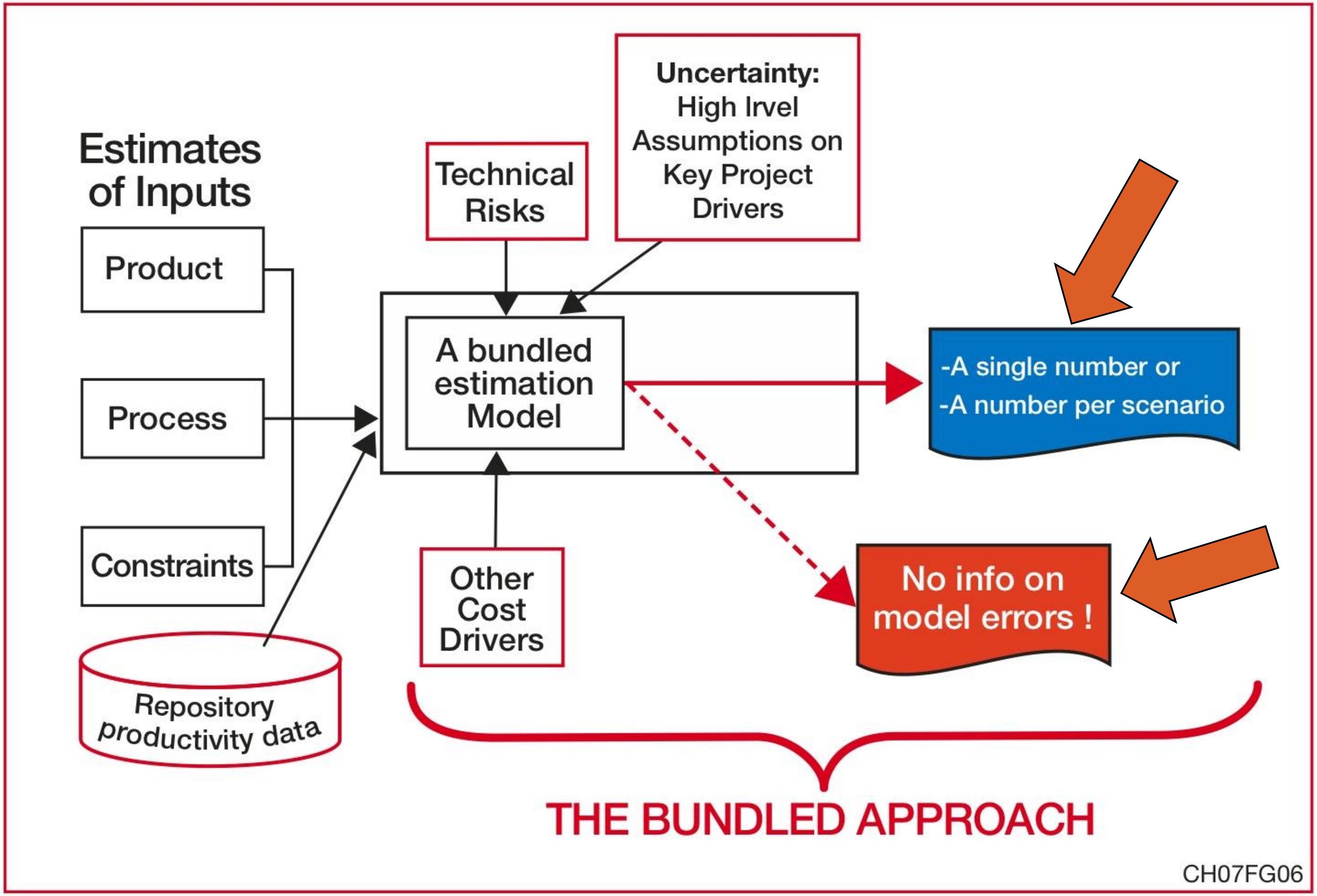


Figure 7.6 Estimation - The Bundle Approach.

Estimation Maths status: The search for gold!

KEMERER 1987

Another Estimation Model:

- With complex mathematical formula
- Claims of being based on +4,000 projects

➢ **Still being marketed in 2014**
...at a very high cost!

KEMERER 1987 on this other estimation model

Small scale replication study – 17 projects

MMRE = 772%

With both large + & -

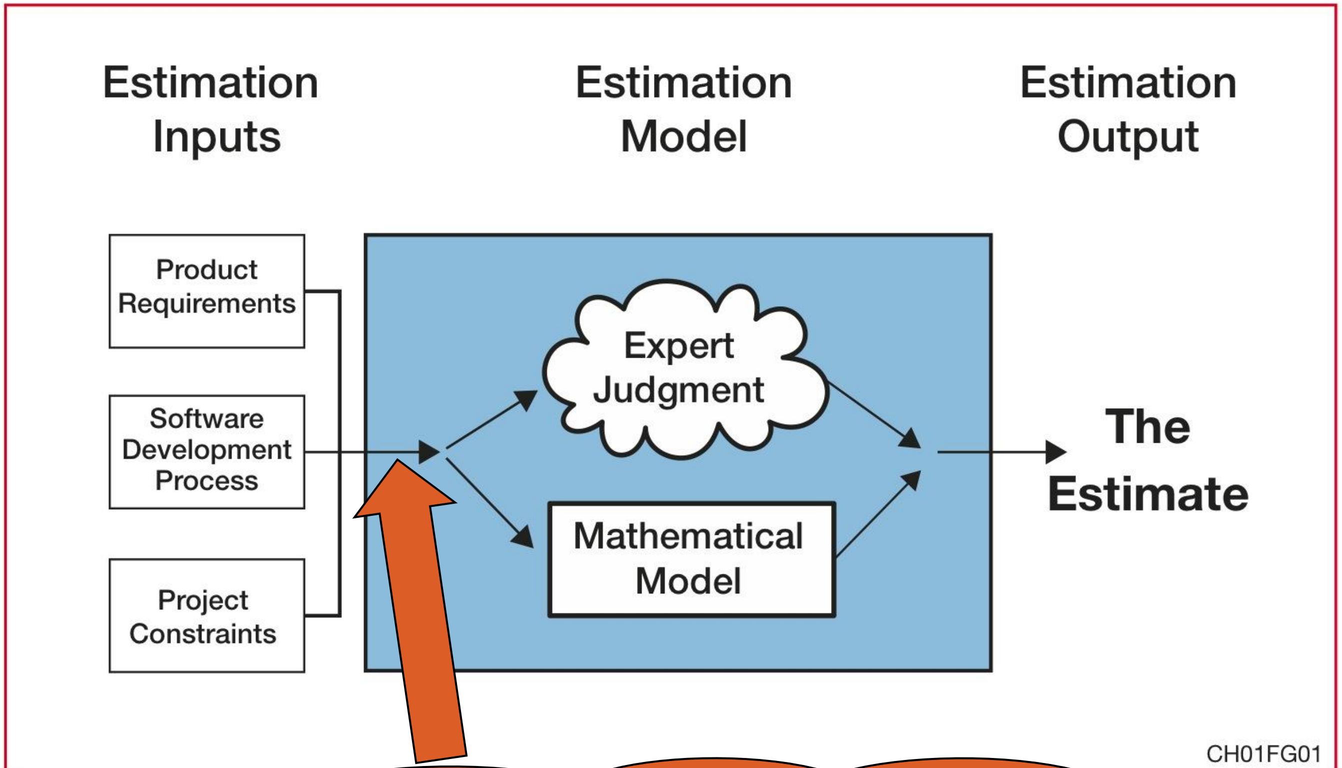
(i.e. cannot be calibrated!)

Larger scale replication study - MMRE

Programming language, size range [in Function Points]	(1) Vendor's black-box estimation tool (%)	(2) White-box models built directly from the data (%)		
Access [200,800]	341	15		
C [200, 800]	1653	50		
C++ [70, 500]	97	86		
C++ [750, 1250]	95	24		
Cobol [60, 400]	400	42		
Cobol [401, 3500]	348	51		
Cobol II [80, 180]	89	29		
Cobol II [180, 500]	109	46		
Natural [20, 620]	243	50		
Natural [621, 3500]	347	35		
Oracle [100, 2000]	319	120		
PL1 [80, 450]	274	45		
PL1 [550, 2550]	895	21		
Powerbuilder [60, 400]	95	29		
SQL [280, 800]	136	81		
SQL [801, 4500]	127	45		
Telon [70, 650]	100	22		
Visual Basic [30, 600]	122	54		
Min	89	15		
Max	1,653	120		

Larger scale replication study - MMRE

Programming language, size range [in Function Points]	(1) Vendor's black-box estimation tool (%)	(2) White-box models built directly from the data (%)		
Access [200,800]	341	15		
C [200, 800]	1653	50		
C++ [70, 500]	97	86		
C++ [750, 1250]	95	24		
Cobol [60, 400]	400	42		
Cobol [401, 3500]	348	51		
Cobol II [80, 180]	89	29		
Cobol II [180, 500]	109	46		
Natural [20, 620]	243	50		
Natural [621, 3500]	347	35		
Oracle [100, 2000]	319	120		
PL1 [80, 450]	274	45		
PL1 [550, 2550]	895	21		
Powerbuilder [60, 400]	95	29		
SQL [280, 800]	136	81		
SQL [801, 4500]	127	45		
Telon [70, 650]	100	22		
Visual Basic [30, 600]	122	54		
Min	89	15		
Max	1,653	120		



Another Key Issue:
The quality of the inputs to the estimation models

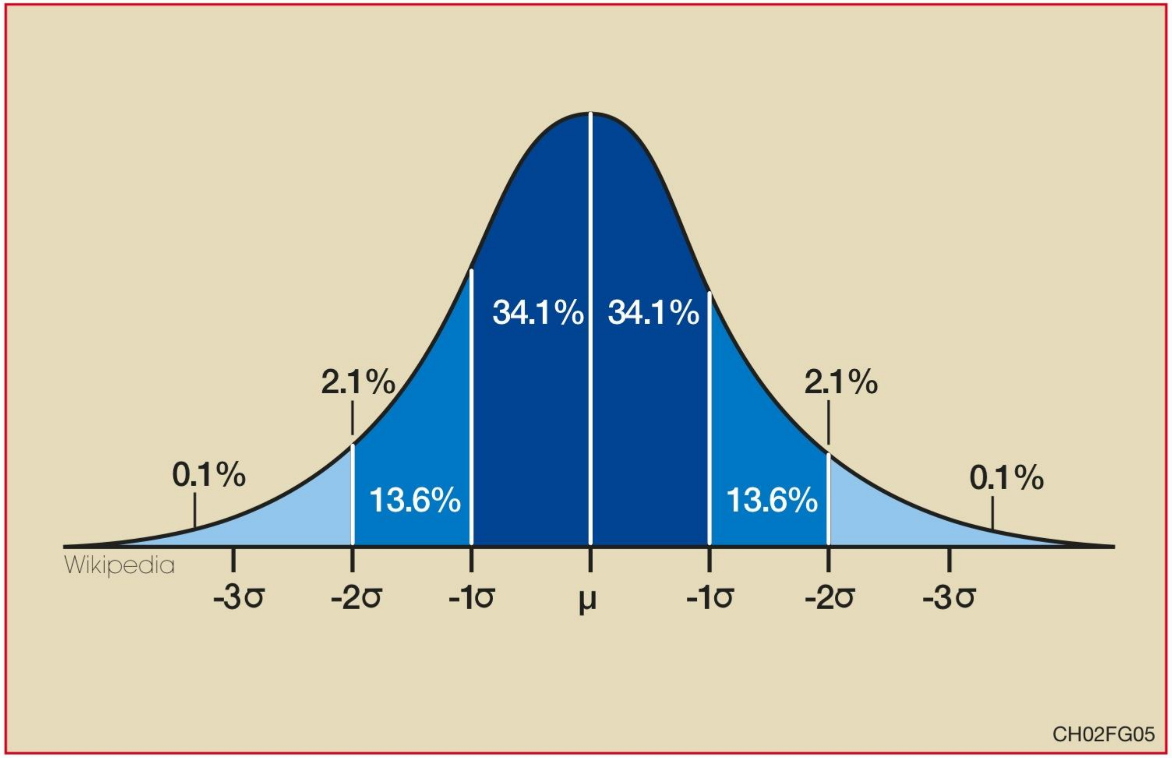


Figure 2.5 A Normal distribution and the standards deviations.

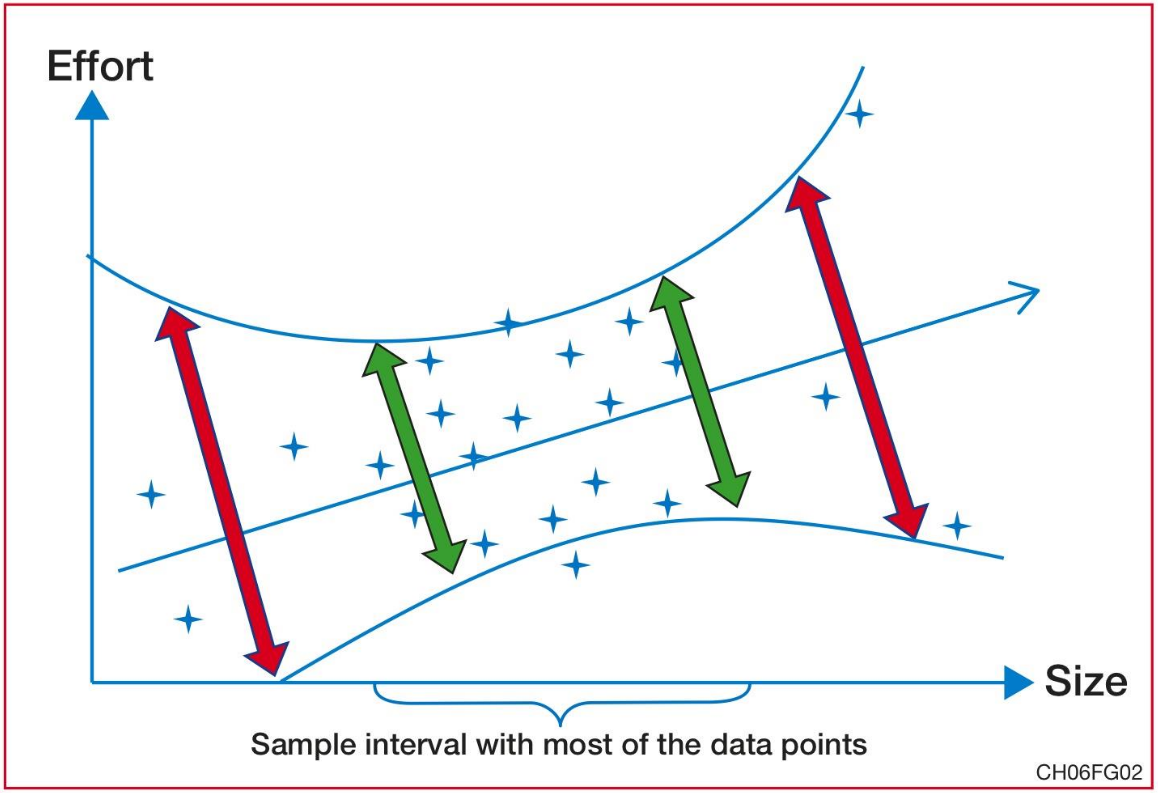


Figure 6.2 Confidence Intervals & Sample Intervals.

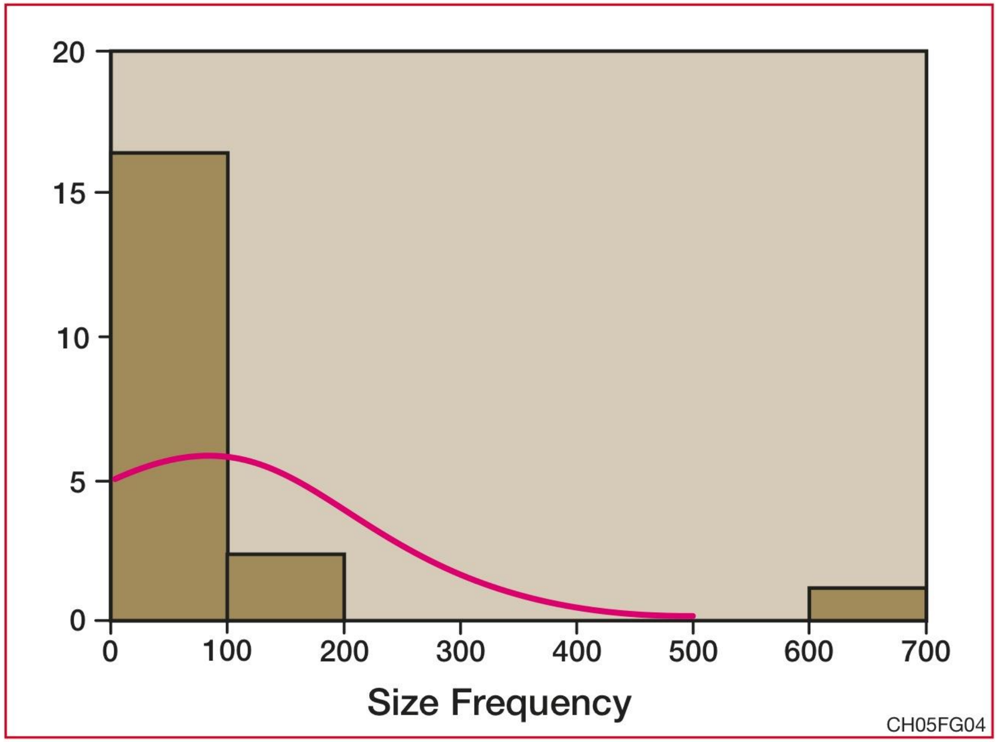
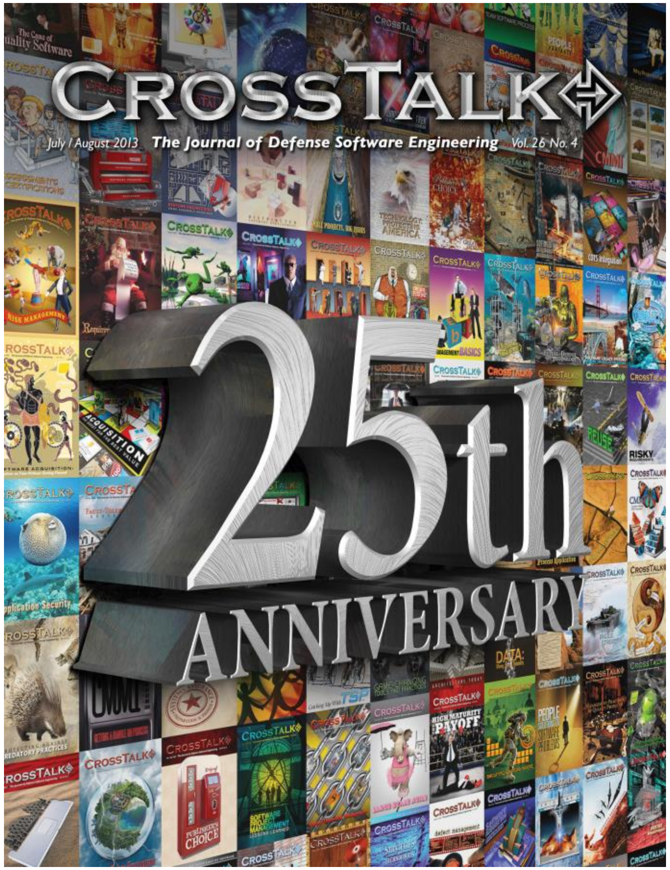
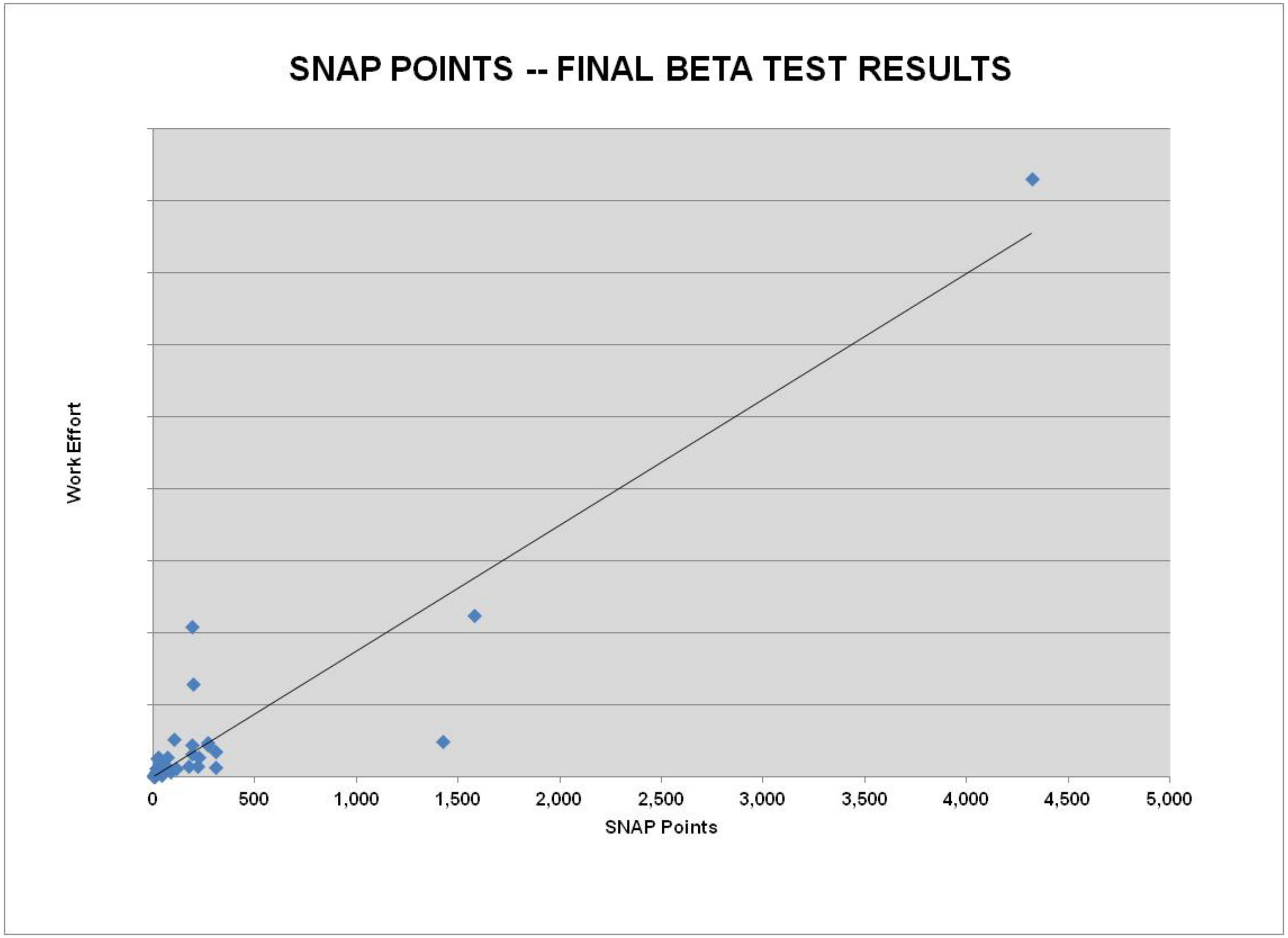


Figure 5.4 Frequency distribution of the size (independent variable) in Table 5.1 with N = 212.

A New Software
Metric to Complement
Function Points
The Software Non-functional Assessment Process (SNAP)





Author → Very strong relationship of SNAP with Effort

→ $R^2 = 0,89$ (R^2 max = 1,0)

Author's assertion on *Figure 4:*

- R^2 = .89 Significance F = 1.7 * 10-23 Spearman = .85 Runs = pass

&

- Spearman test for rank correlation of .85, with an associated confidence of statistical significance of greater than 99% (p-value <.0001).

Author’s assertion on *Figure 4:*

- R^2 = .89 Significance F = 1.7 * 10-23 Spearman = .85 Runs = pass

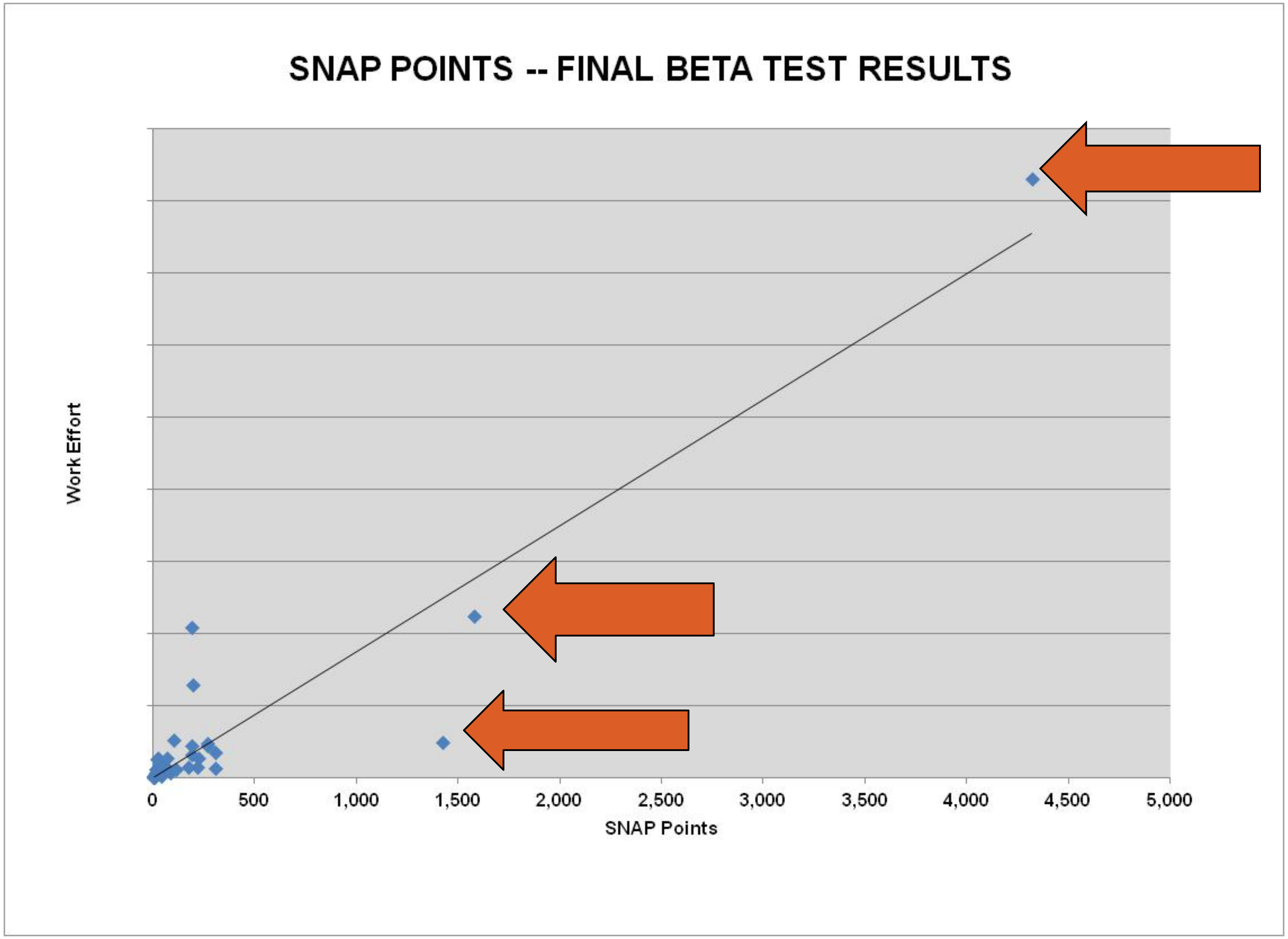
&

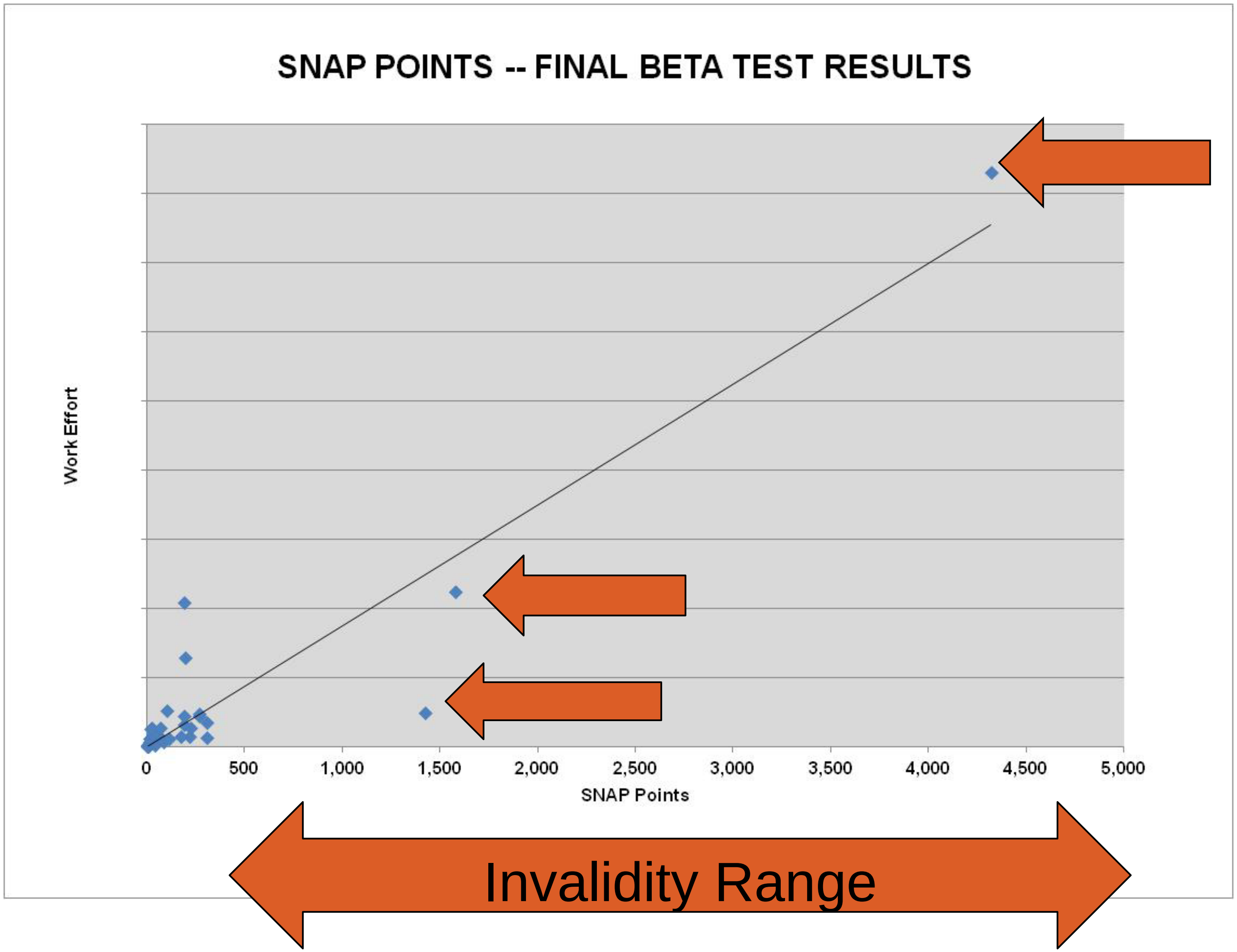
- Spearman test for rank correlation of .85, with an associated confidence of statistical significance of greater than 99% (p-value <.0001).

But:

These numbers and stats are **invalid**

the necessary requiremenst for a regression are not met!
Presence of large outliers which distorts all stats numbers
Meaningless!!


SNAP POINTS -- FINAL BETA TEST RESULTS
Work Effort
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
SNAP Points


SNAP POINTS -- FINAL BETA TEST RESULTS
Work Effort
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
SNAP Points
Invalidity Range

What it really looked like for the range for which there is enough data points

Approxmimatively:
An $R^2 = 0.3$
Not $R^2 = 0.89$
(R^2 max = 1,0)

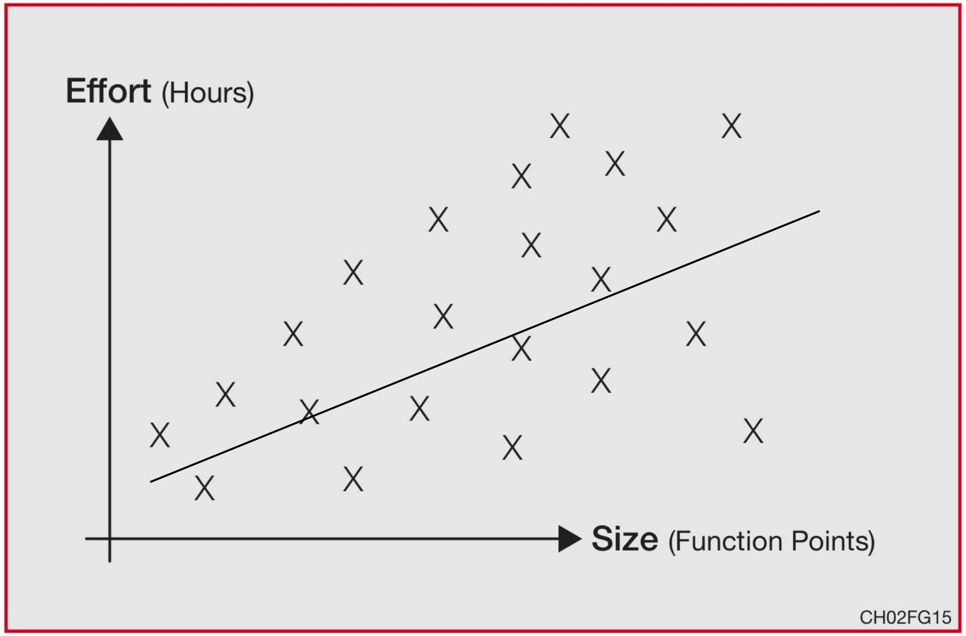
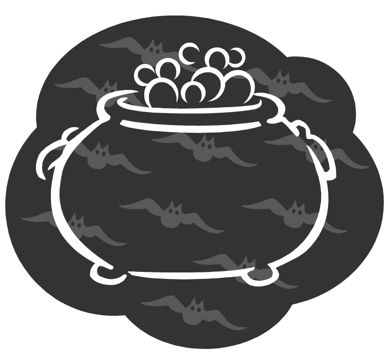
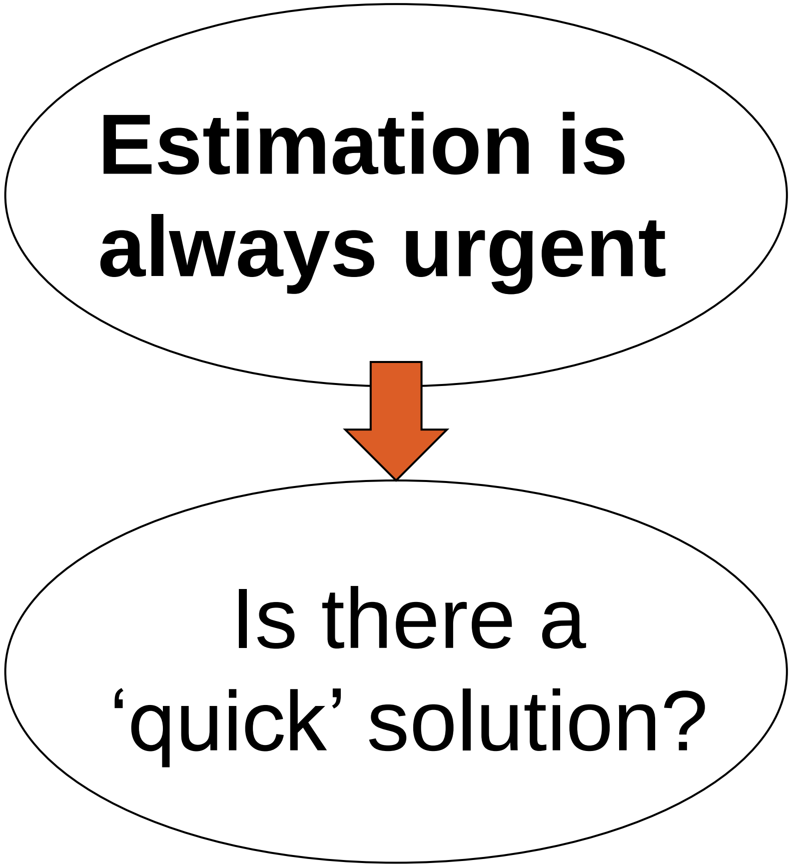


Figure 2.15 Wedge-shaped dataset in software engineering.

CONCLUSION: invalid approach to empirically adopt SNAP!

Hell is paved all over
with good intentions!




Estimation is always urgent
Is there a 'quick' solution?

Estimation is always urgent

Is there a 'quick' solution?

The web!

Estimation is always urgent

Is ‘**quick**’ the right question?

The web!

Estimation is always **urgent**

Isn't it '**quality**' of the estimation model the right question?

The web!

Estimation is always urgent

Isn't '**quality**' of the estimation model the right question?

The web!

Estimation Expected Outcomes

Quick & Easy…

Estimation Outcomes!

Quick & Easy…

COCOMO-like estimation models

The ‘***feel-good***’ → dead end!

→ Quick & Easy…

→

List of topics

1. Estimation: Craft or Engineering?

2. **The estimation phases**

3. Economics concepts for estimation models (fixed-variable costs, economies of scale…)

4. Orphean research issues

Estimation & Uncertainty – Boehm's Cone of Uncertainty

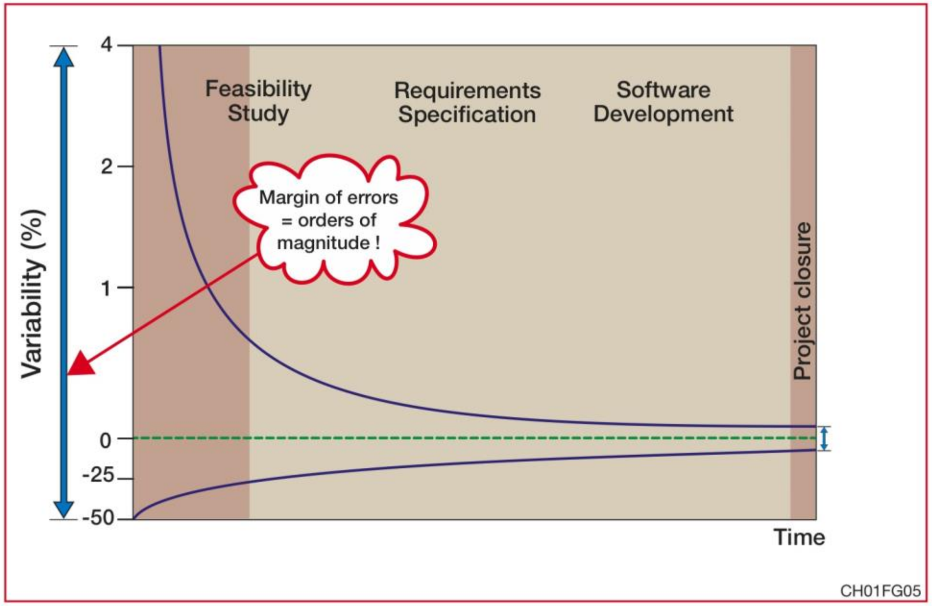


Figure 1.5 Uncertainty decreases over time.

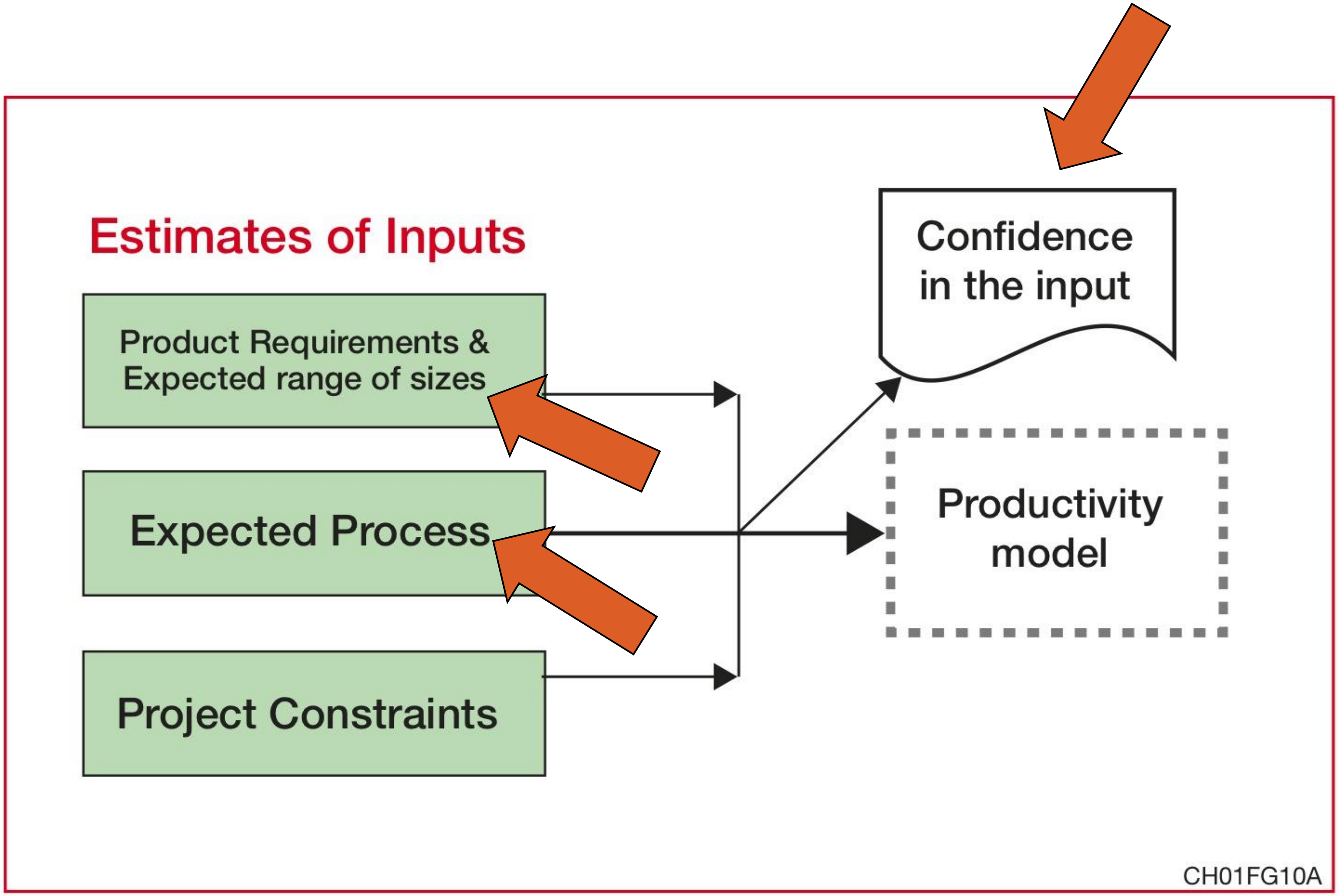


Figure 1.10 Phase A :
Collection of the Inputs for the Estimation Process.

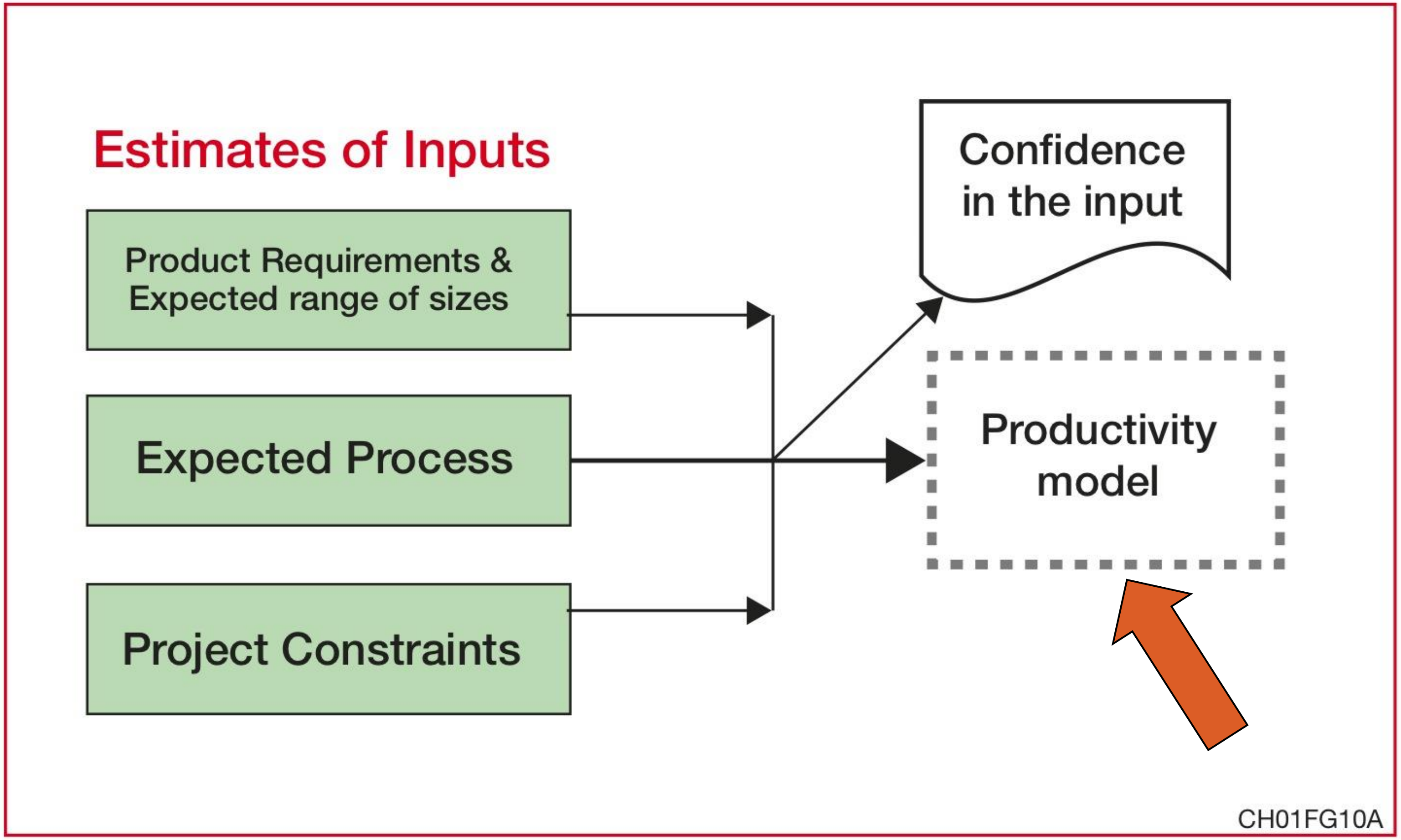


Figure 1.10 Phase A :
Collection of the Inputs for the Estimation Process.

Models Built with completed projects

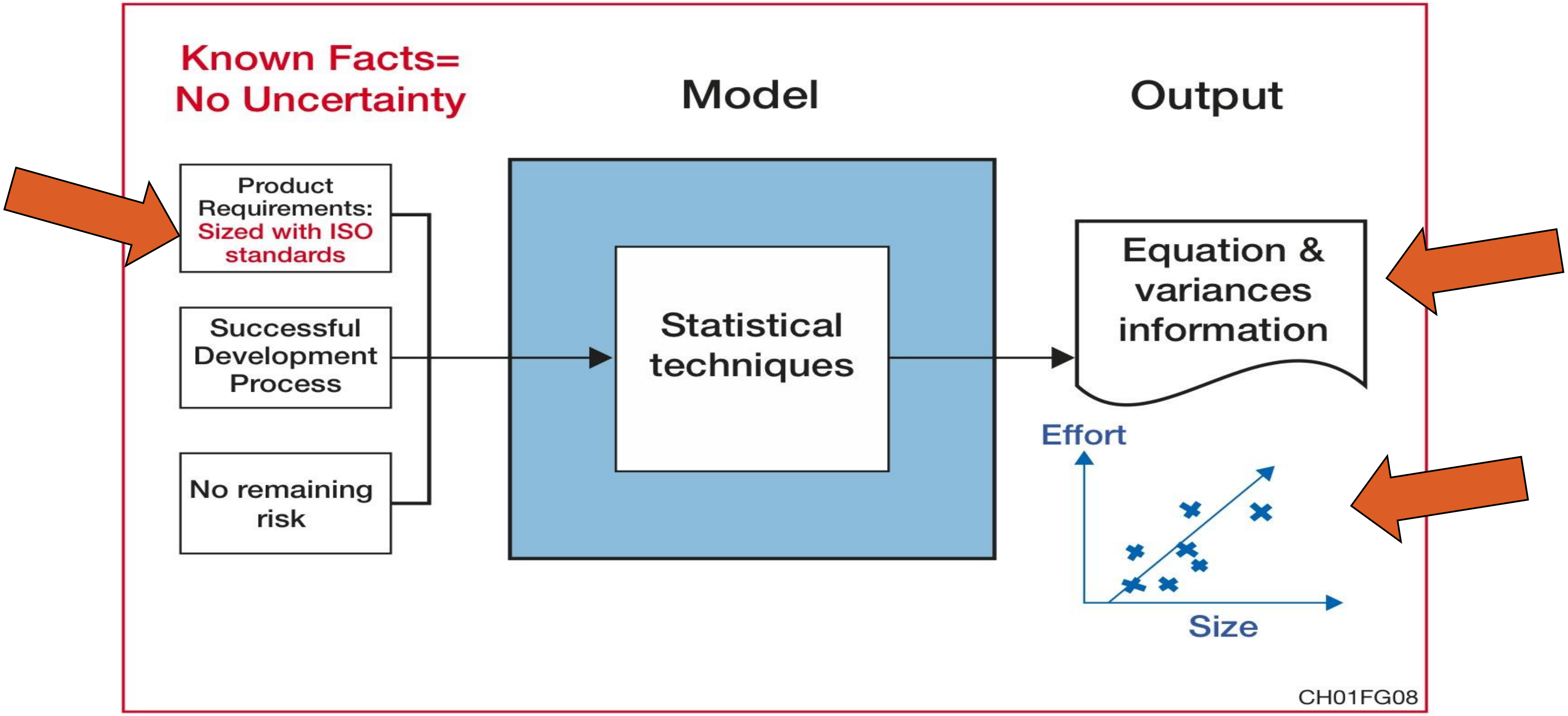
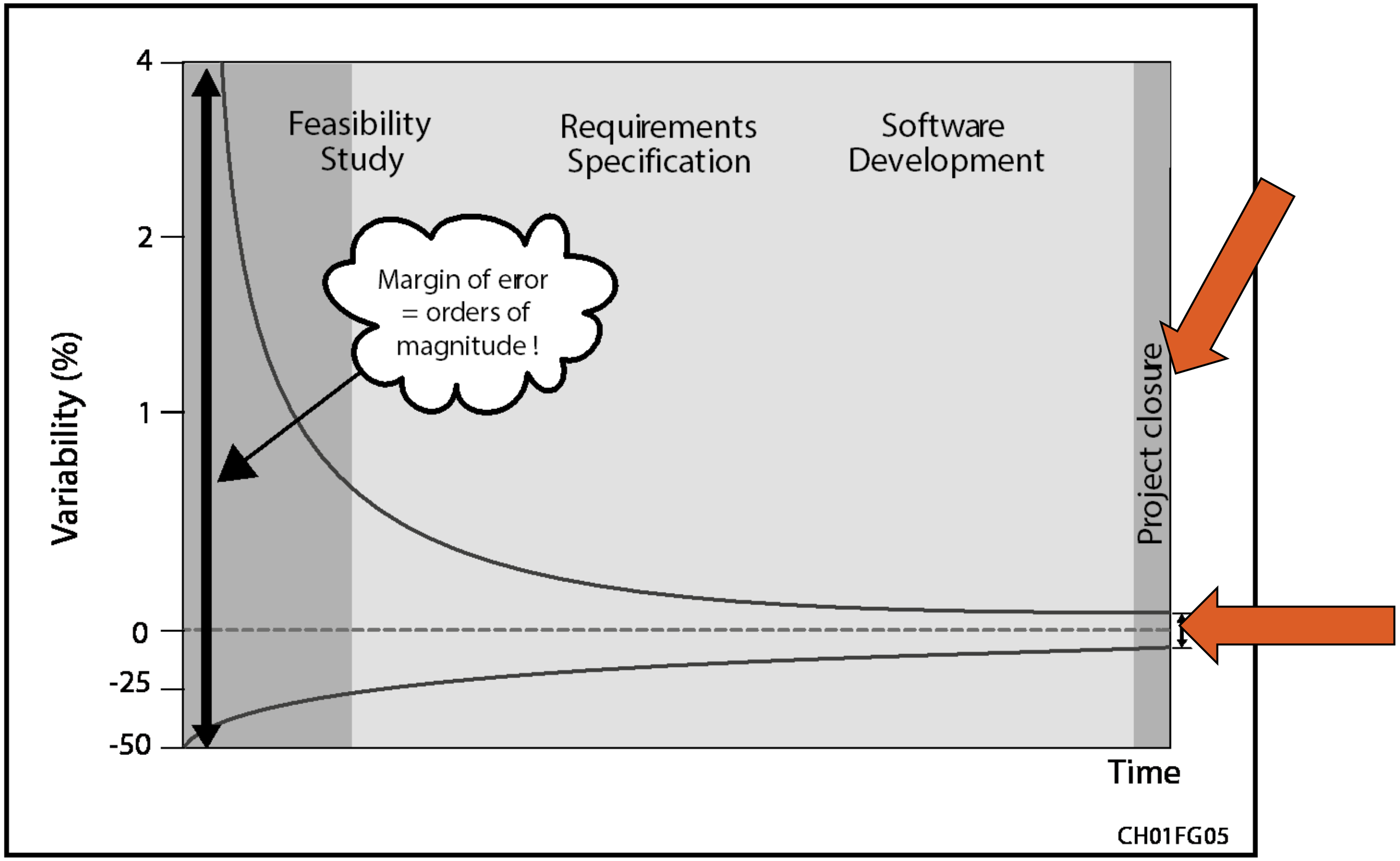
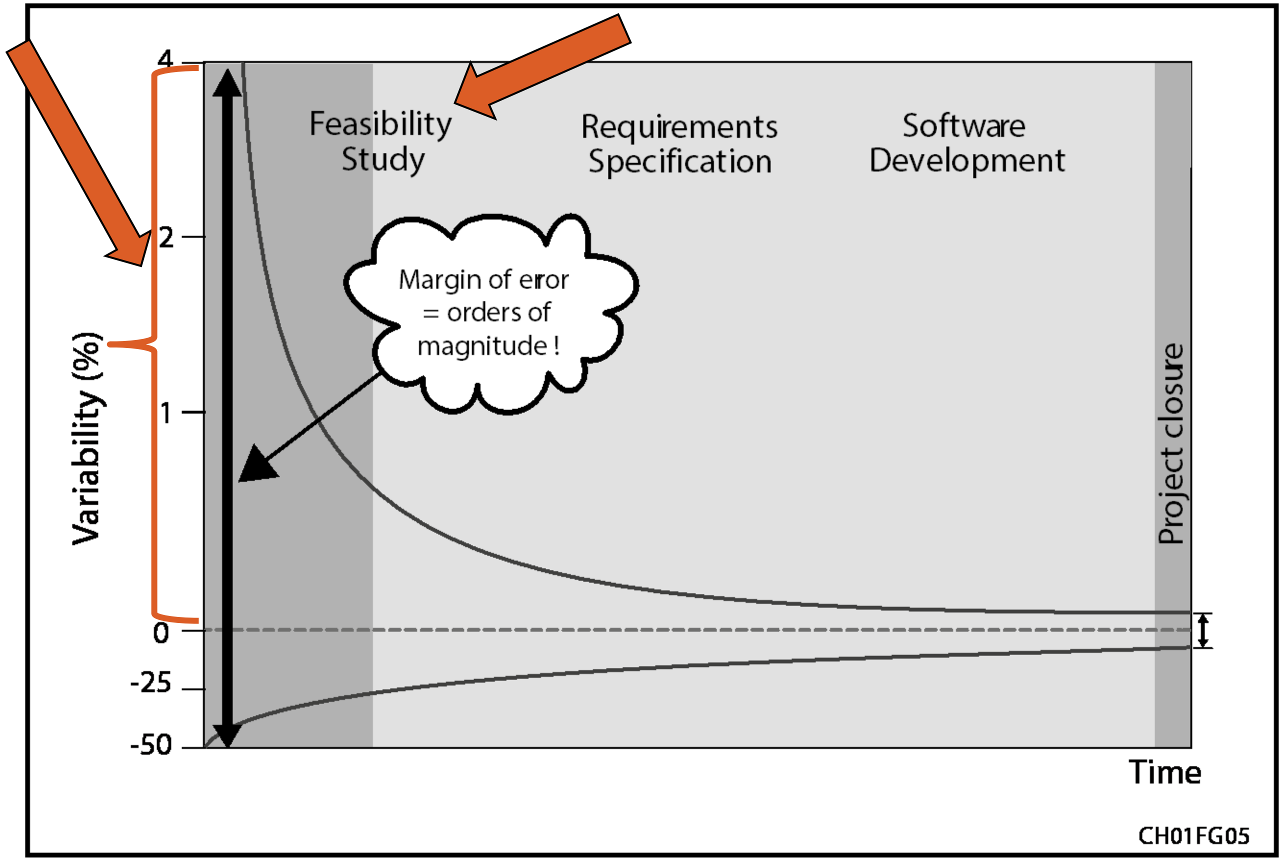


Figure 1.8 The context of a productivity model.

The inputs to productivity models have little uncertainty = Known Facts

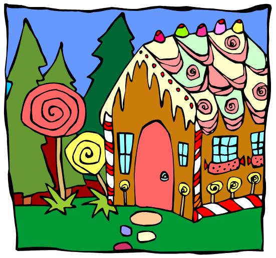


Imprecise Inputs at Feasibility Analysis – Much Greater Error Range



Project Scope = ?

Stakeholders initial wishes



The dreamer

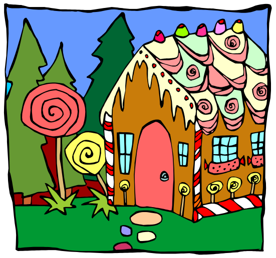
Marketing

The visionnary

Accounting

Project Scope: Detailed & Approved

Stakeholders initial wishes



The dreamer

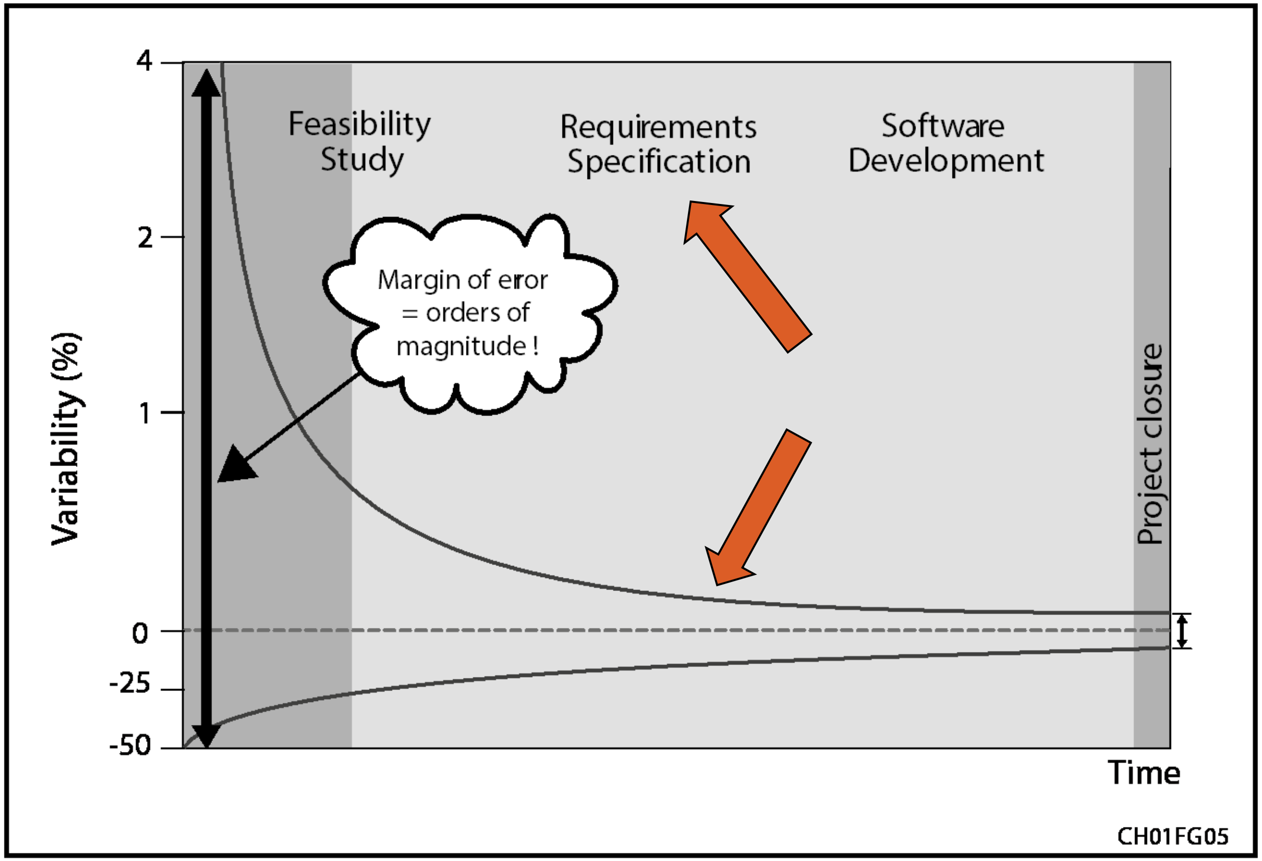
Marketing

The visionnary

Accounting

Agreed Project Scope!

Estimation Models: The Uncertainty Cone: Requirements Specs



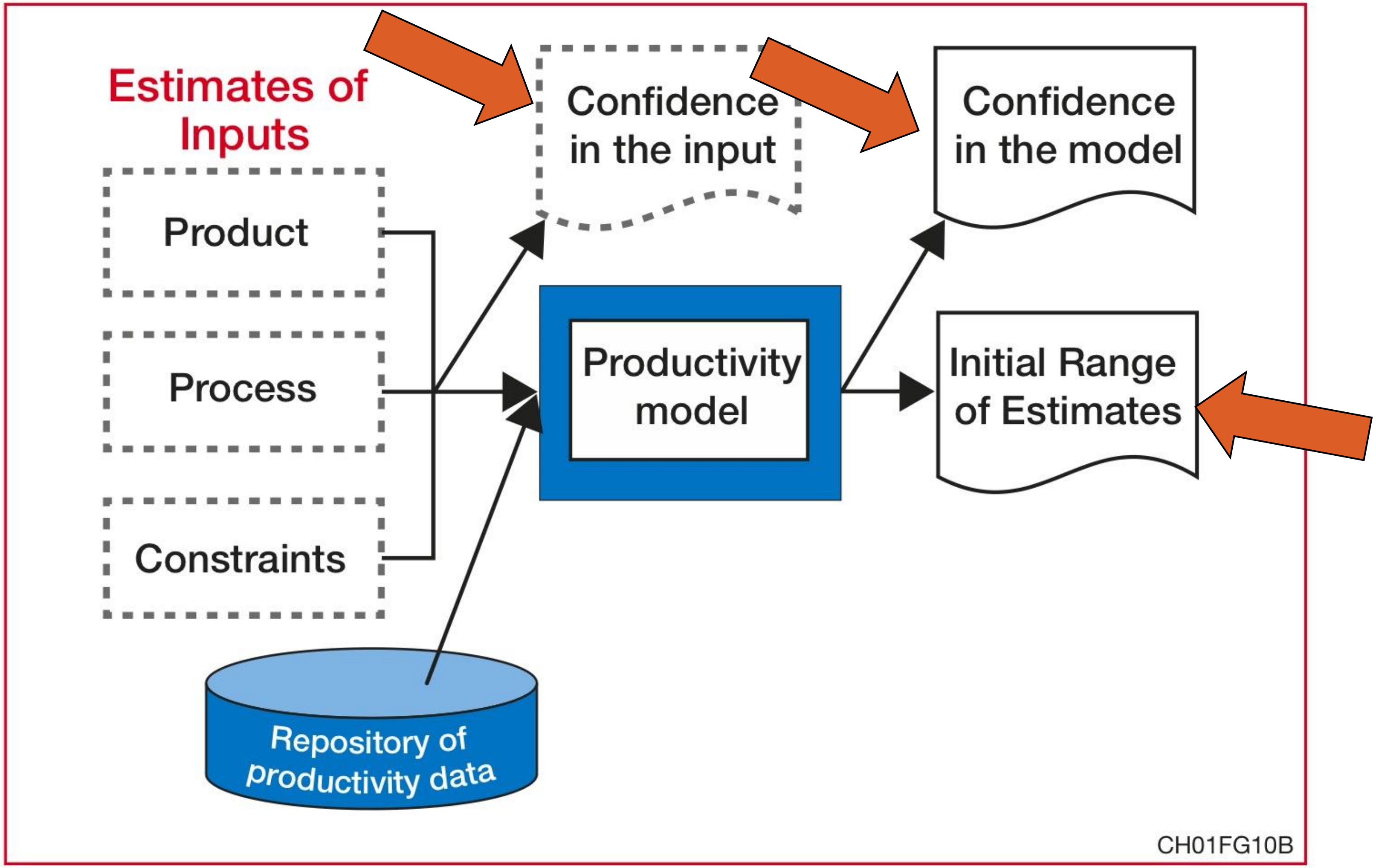


Figure 1.10 Phase B : Execution of the productivity model.

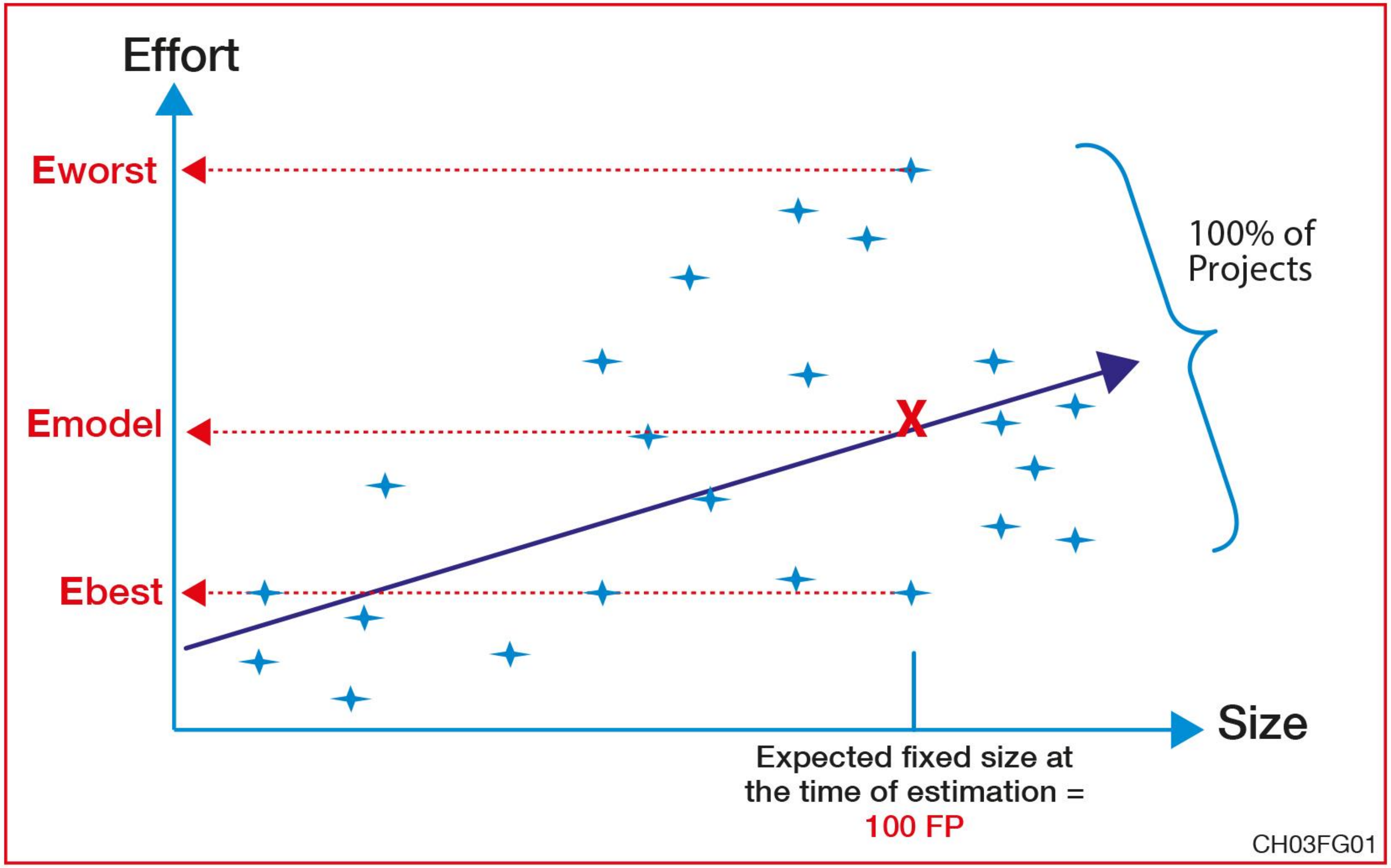


Figure 3.1 Best & Worst Case Scenarios.

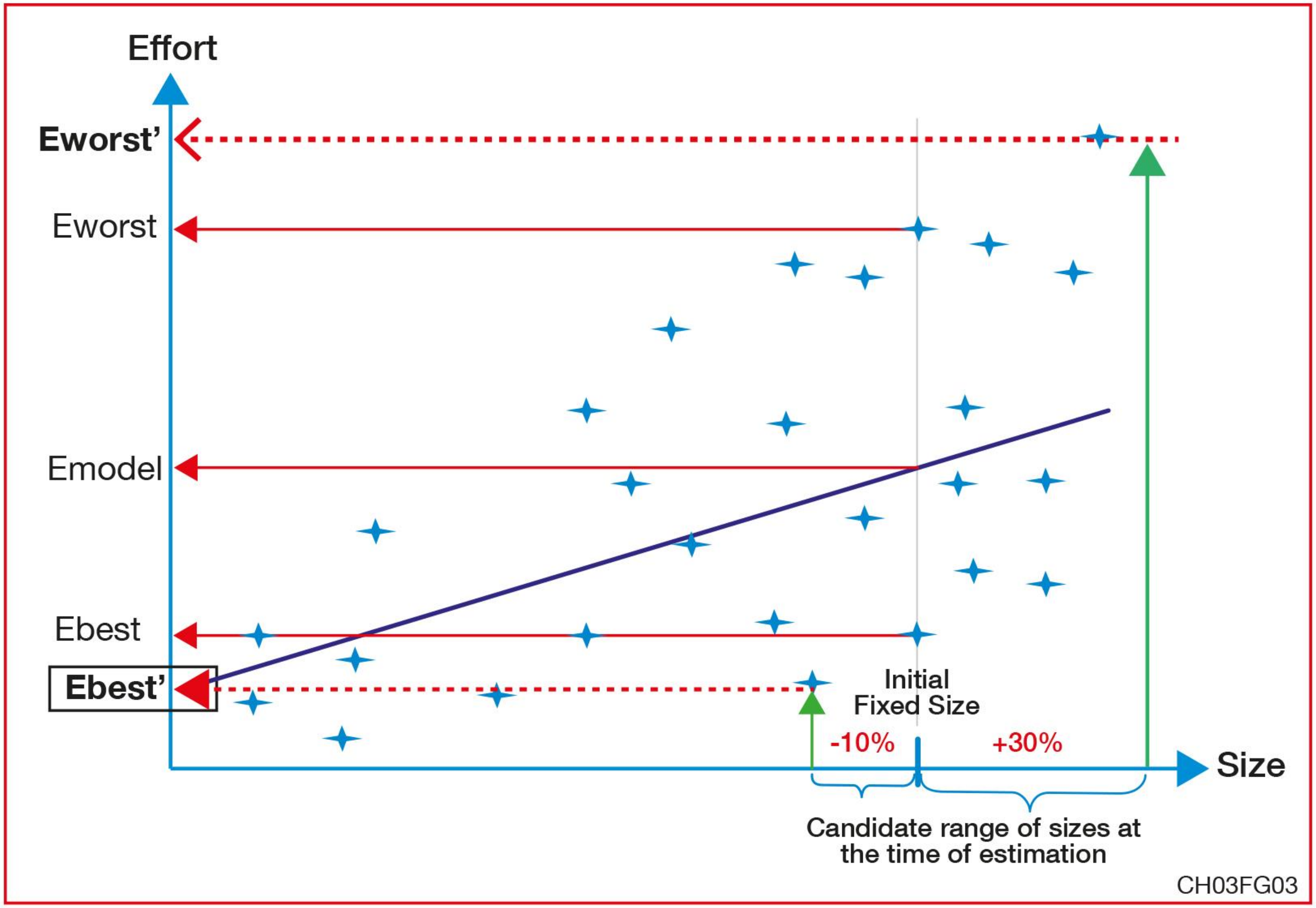


Figure 3.3 Best & Worst Scenarios & Size Uncertainly.

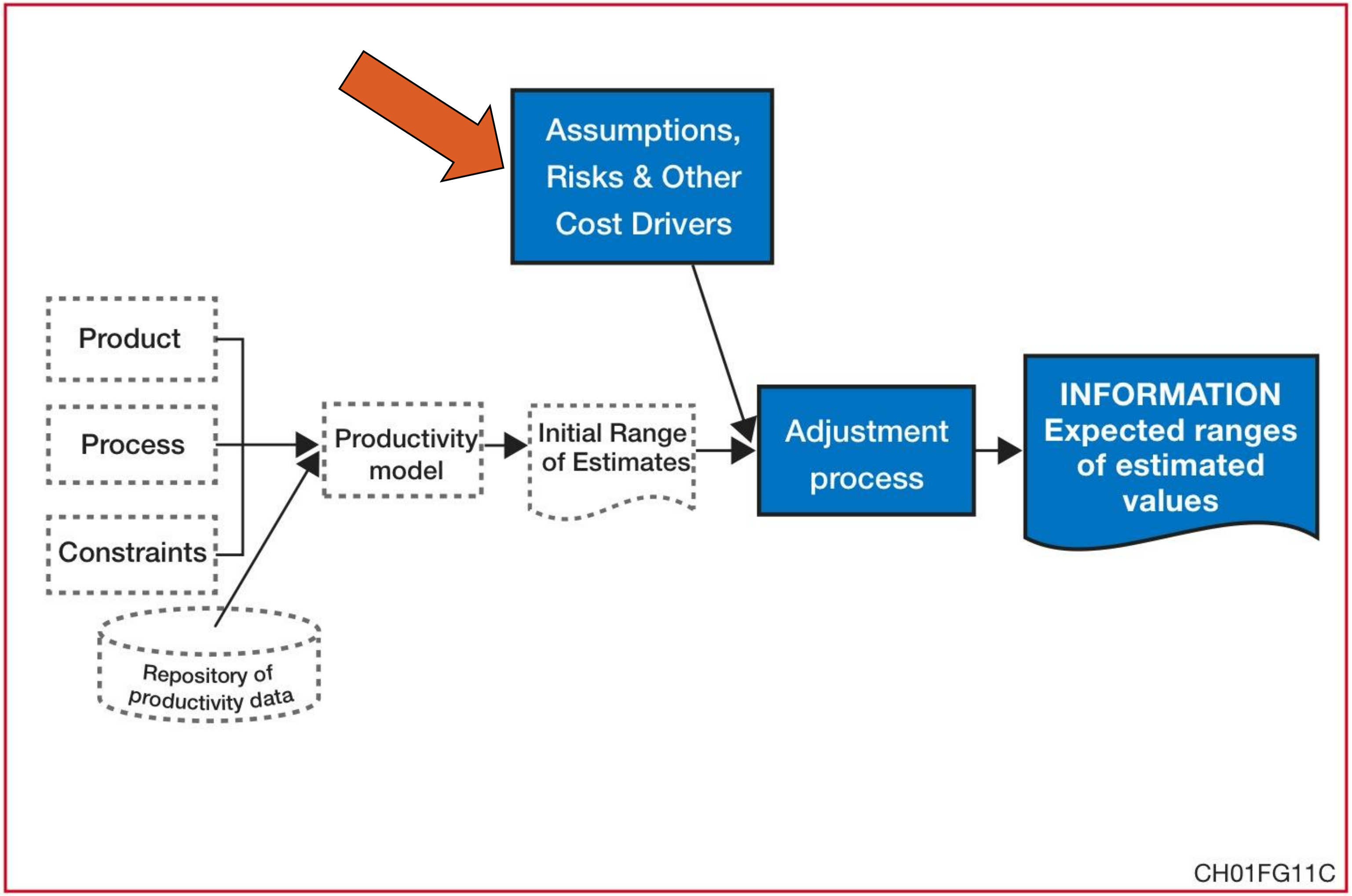


Figure 1.11 Phase C : The adjustment process.

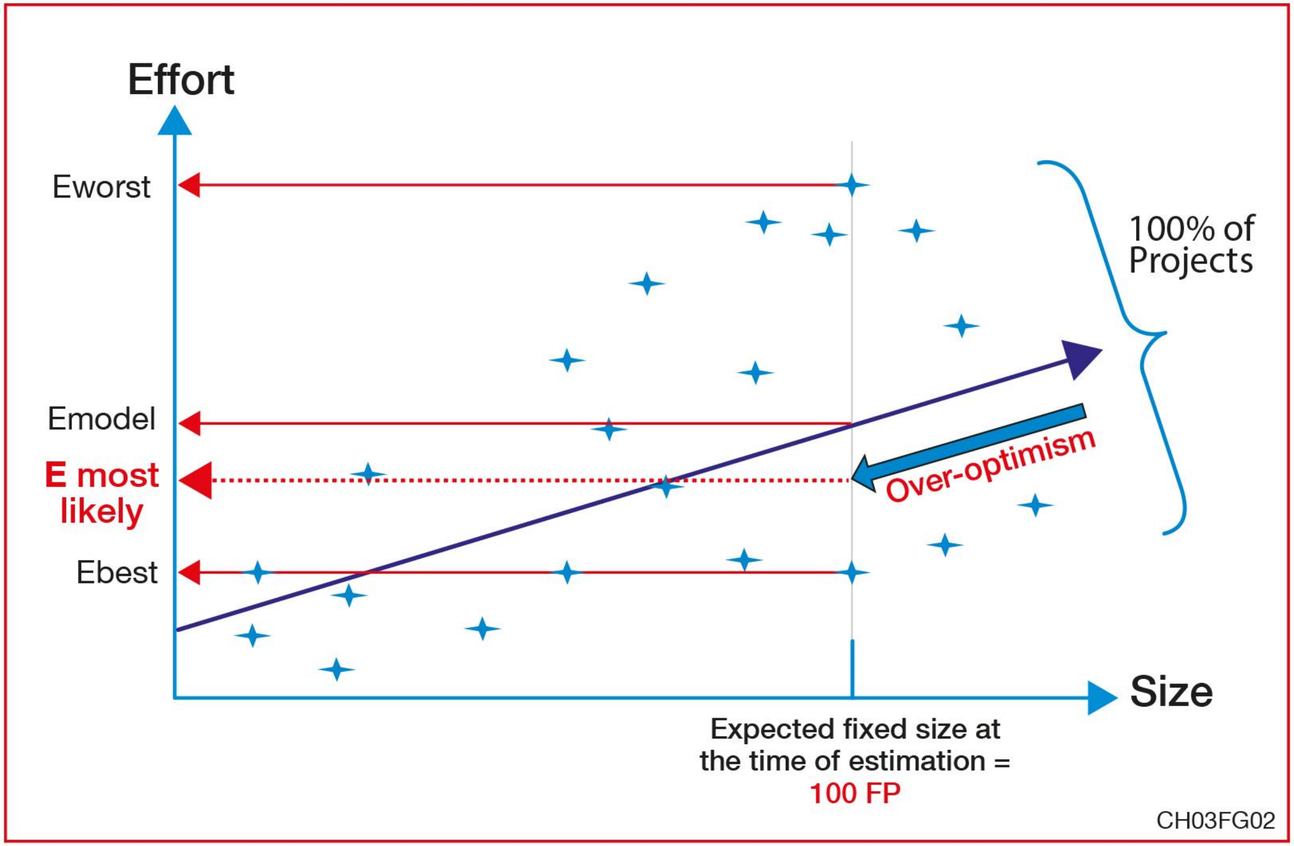


Figure 3.2 Most Likely Scenario & Over-optimism.

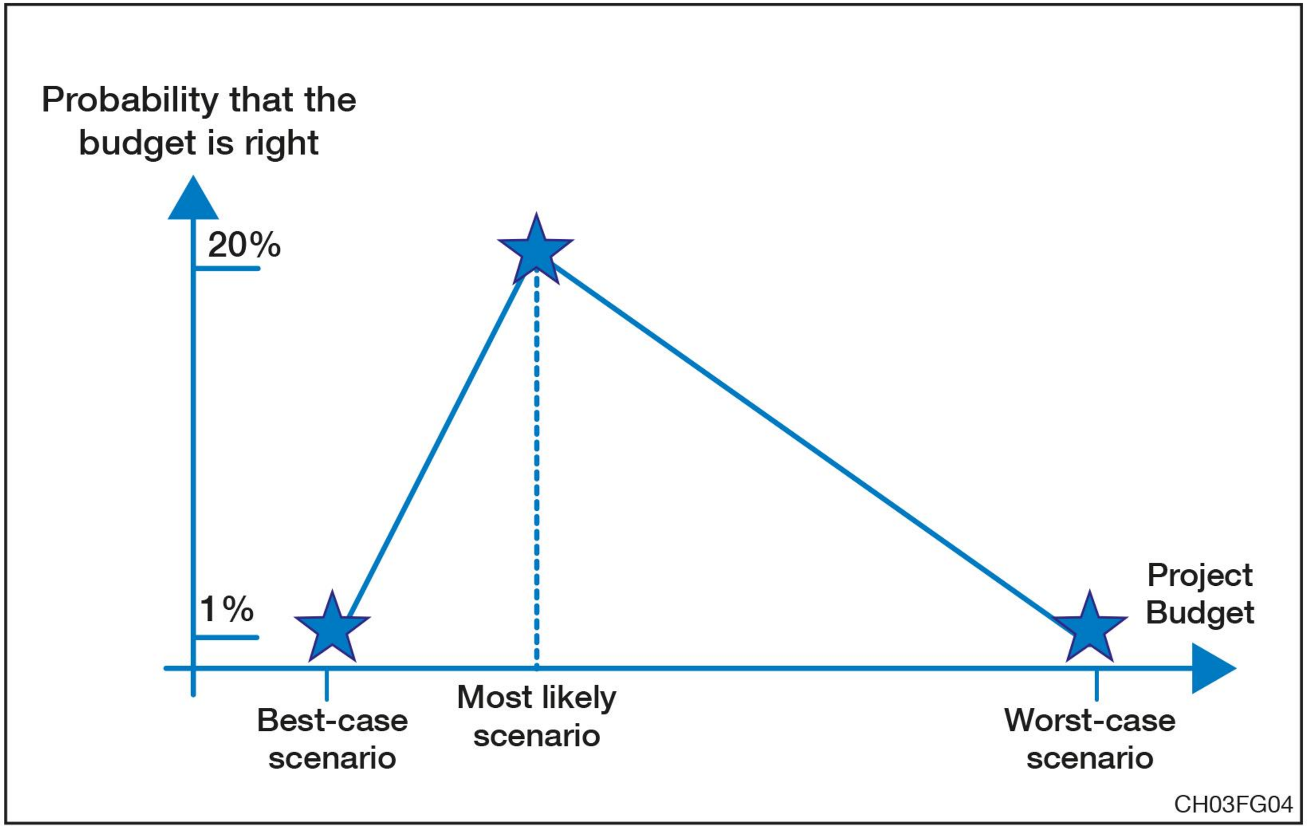


Figure 3.4 Probability Distribution of Scenarios.

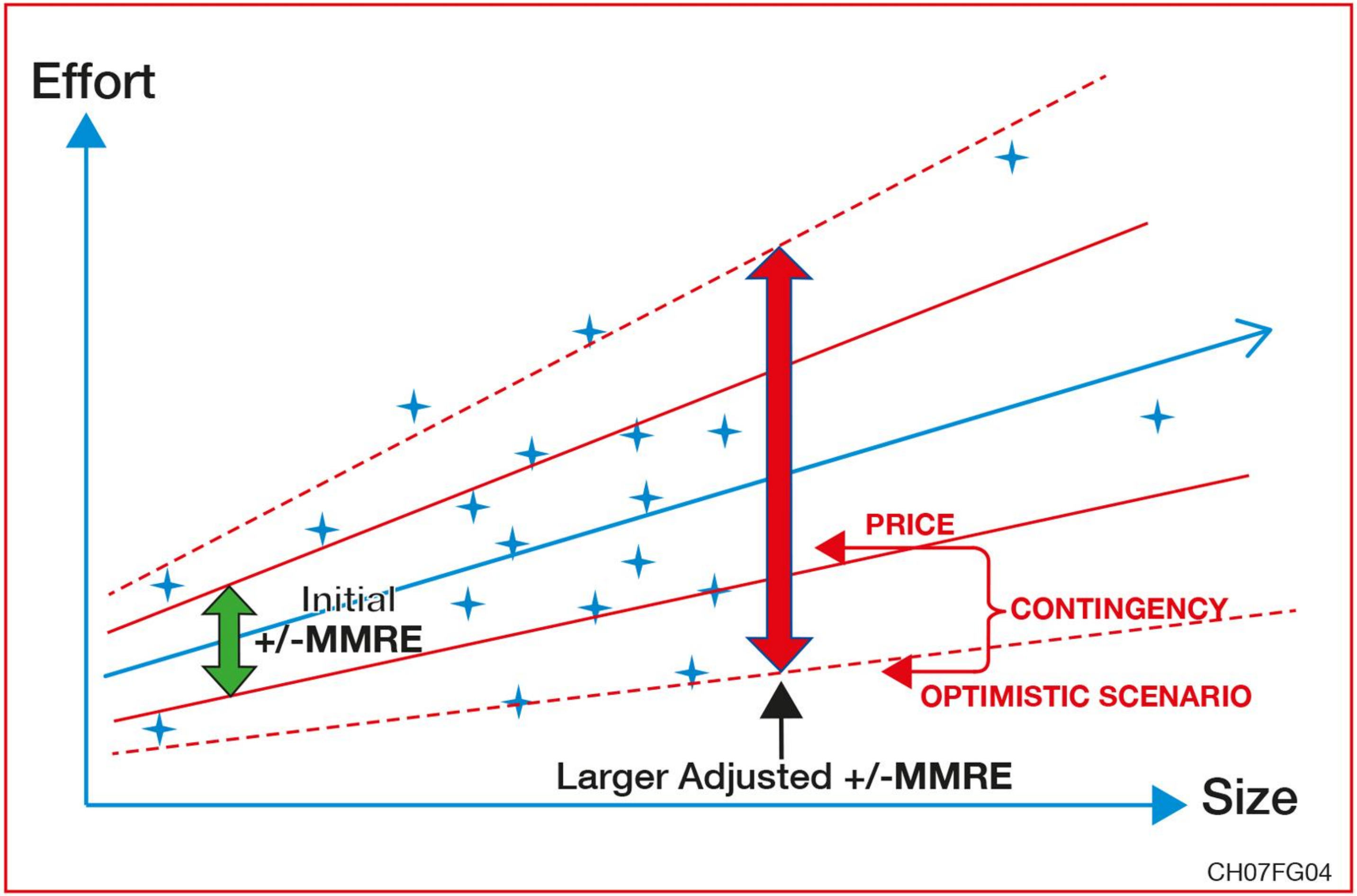


Figure 7.4 Project budget = contingency = price - Optimistic scenario.

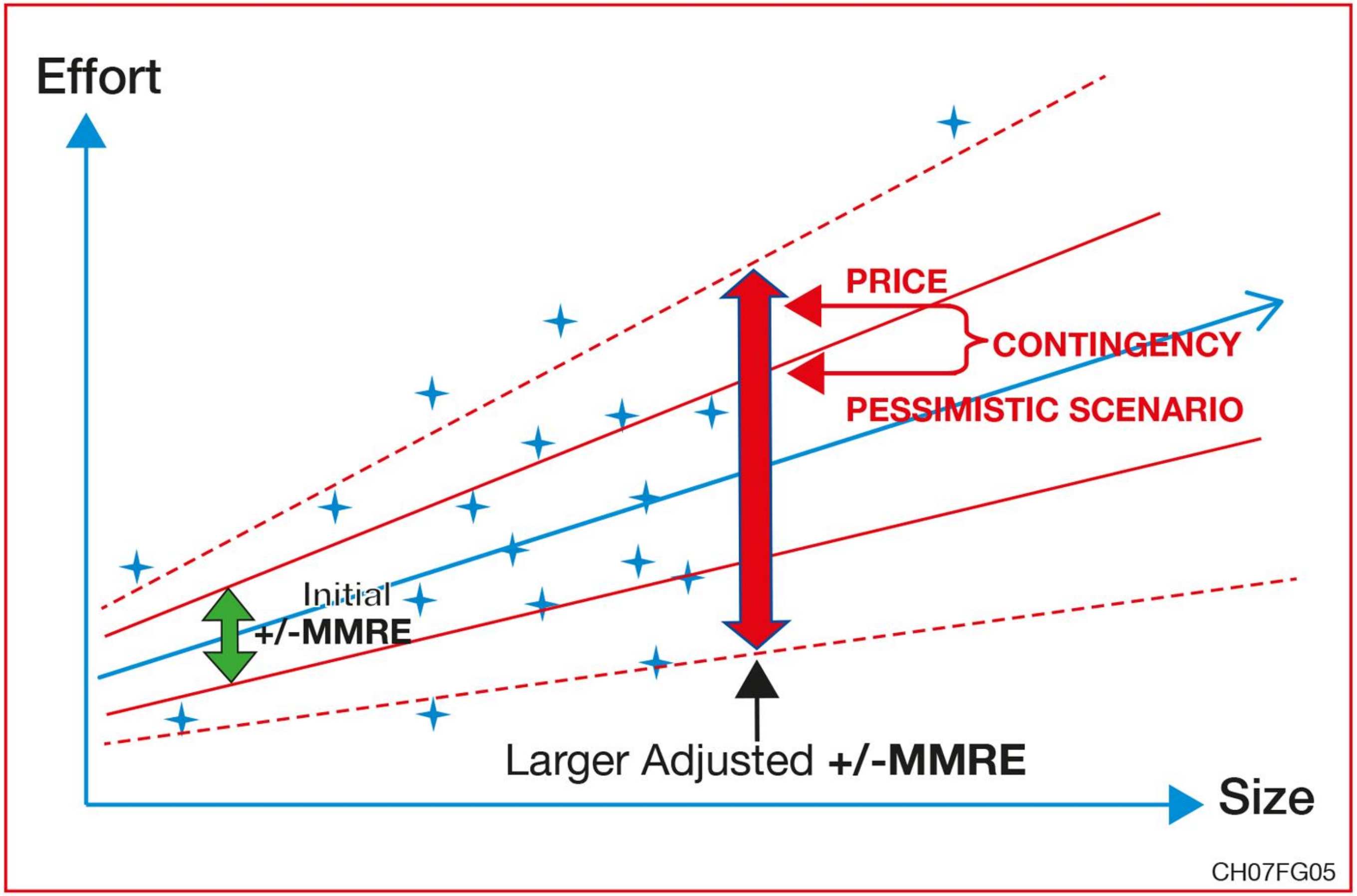


Figure 7.5 Project budget = contingency = price - Pessimistic scenario.

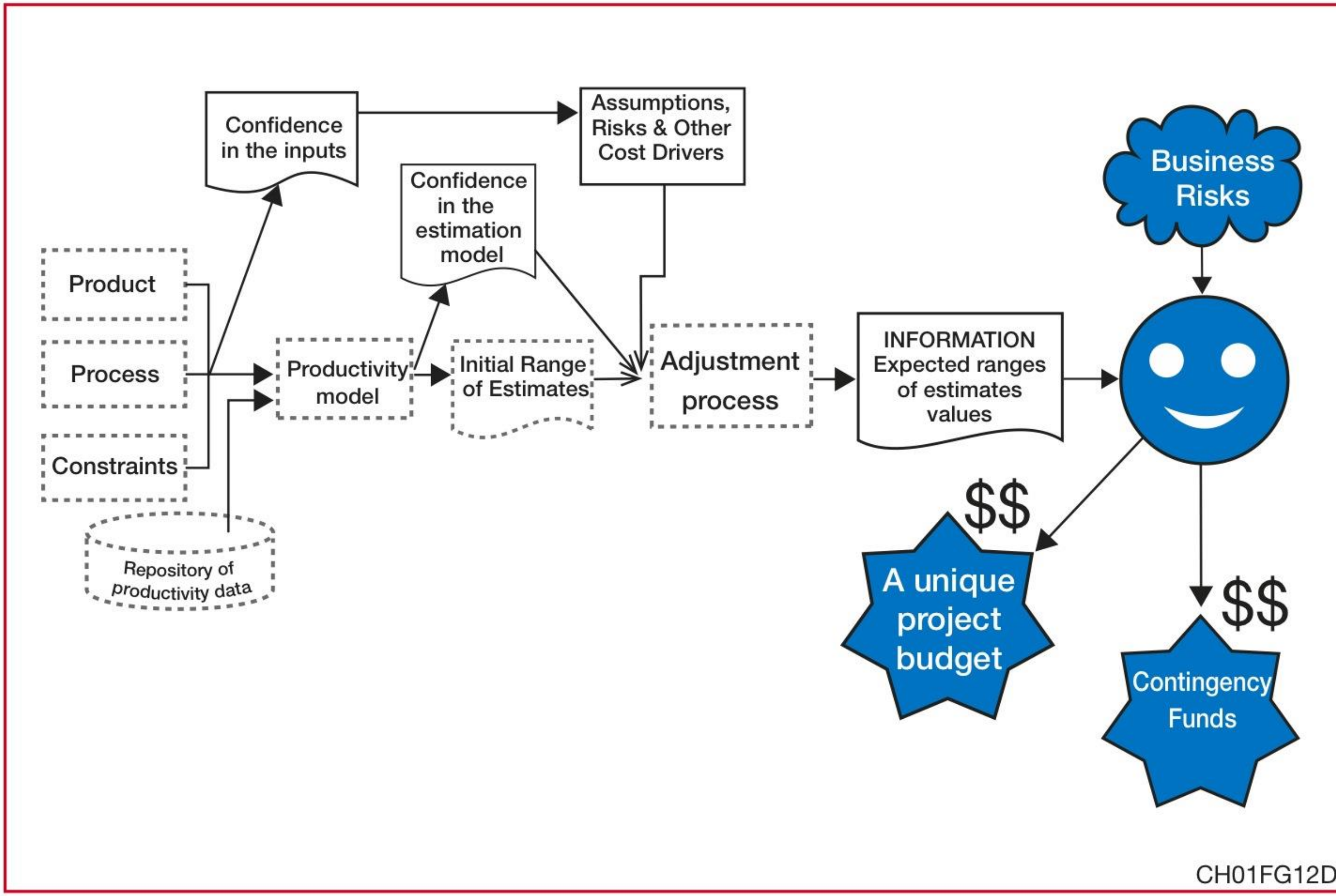


Figure 1.12 Phase D : Budgeting decision.

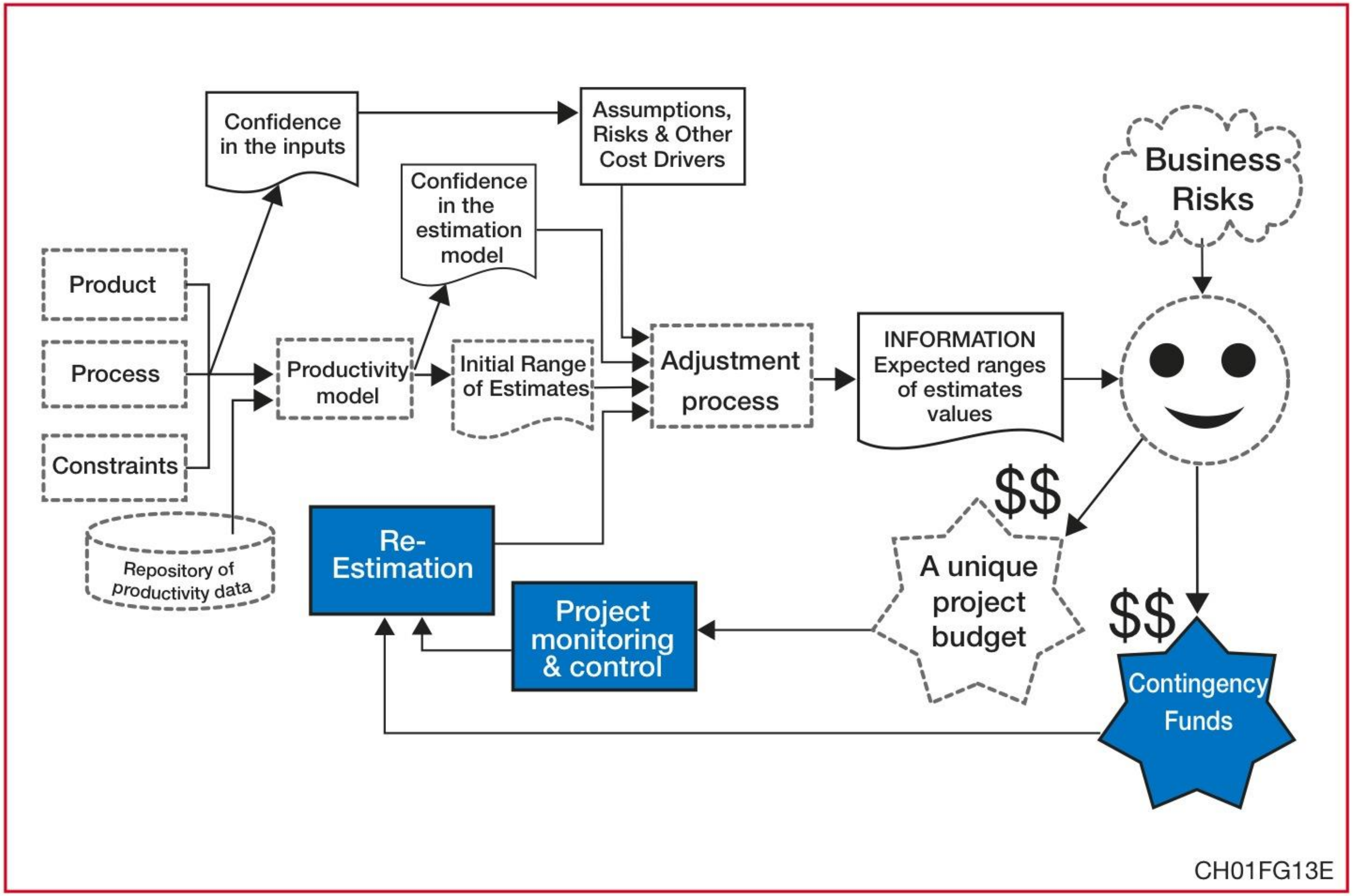


Figure 1.13 Phase E : Re-Estimation.

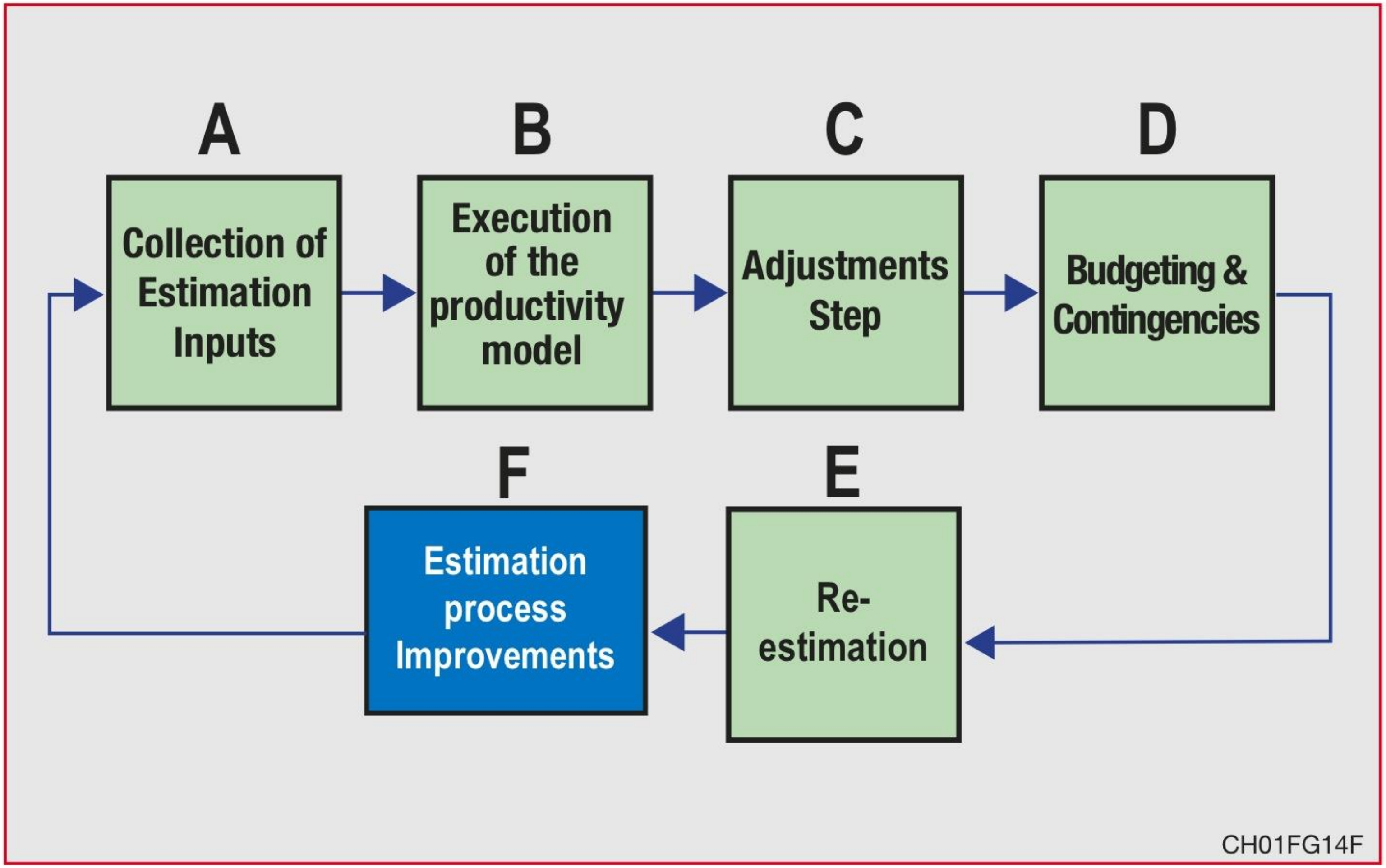
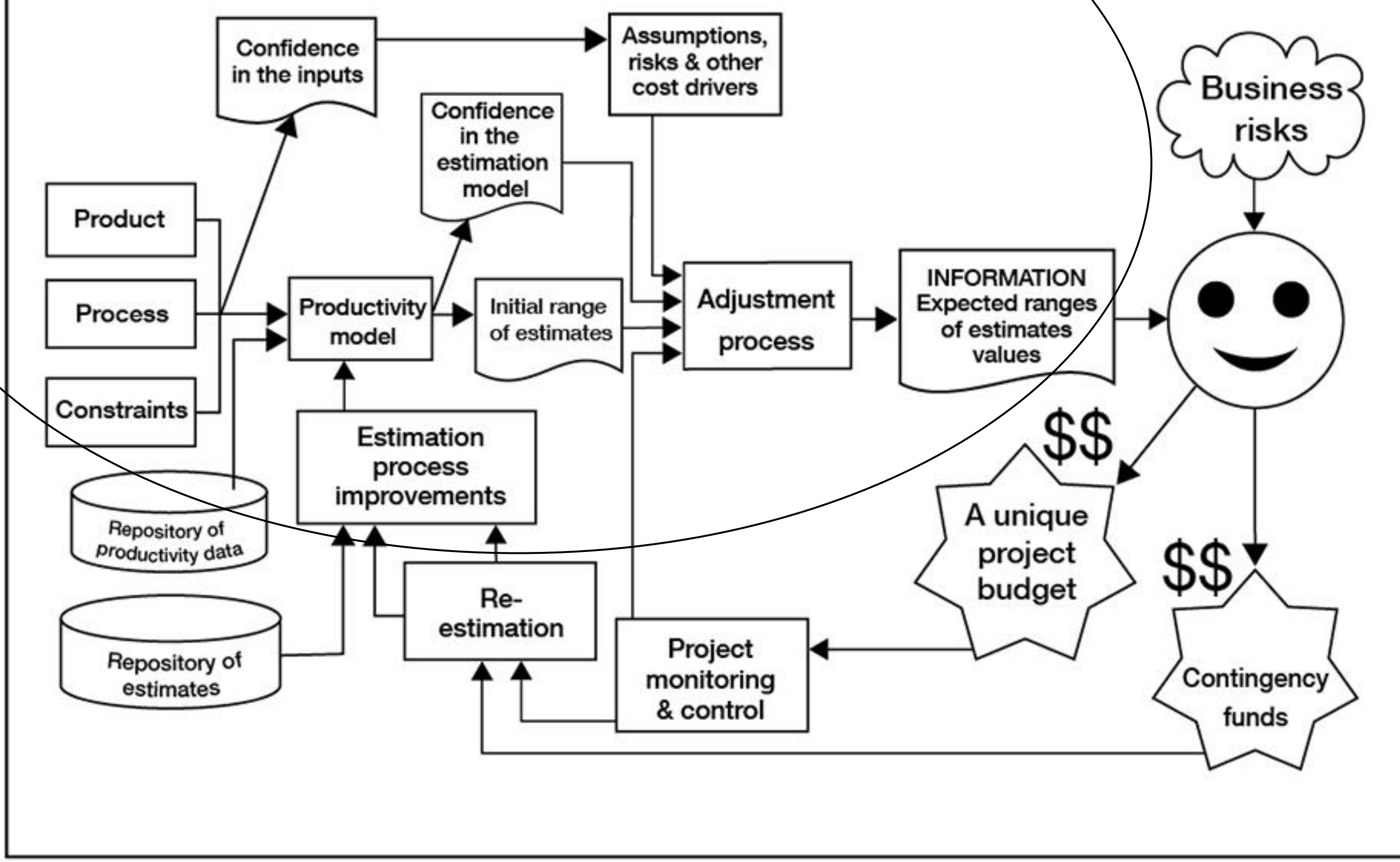
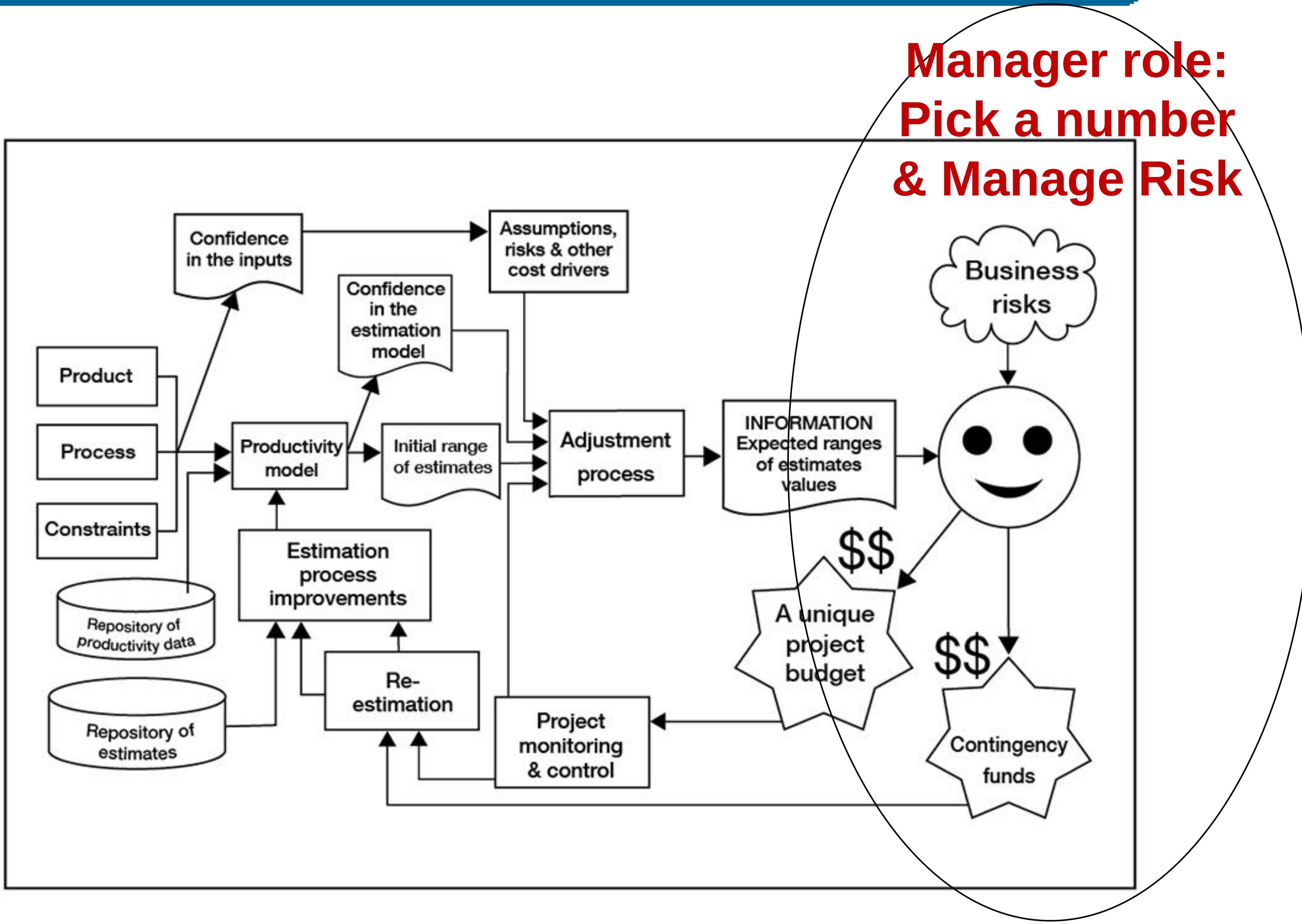


Figure 1.14 Phase F: Estimation Feedback Loop.

Estimator role:
Provide Information



Manager role: Pick a number & Manage Risk



List of topics

1. Estimation: Craft or Engineering?
2. The estimation phases
3. **Economic concepts for estimation models**
4. Orphean research issues

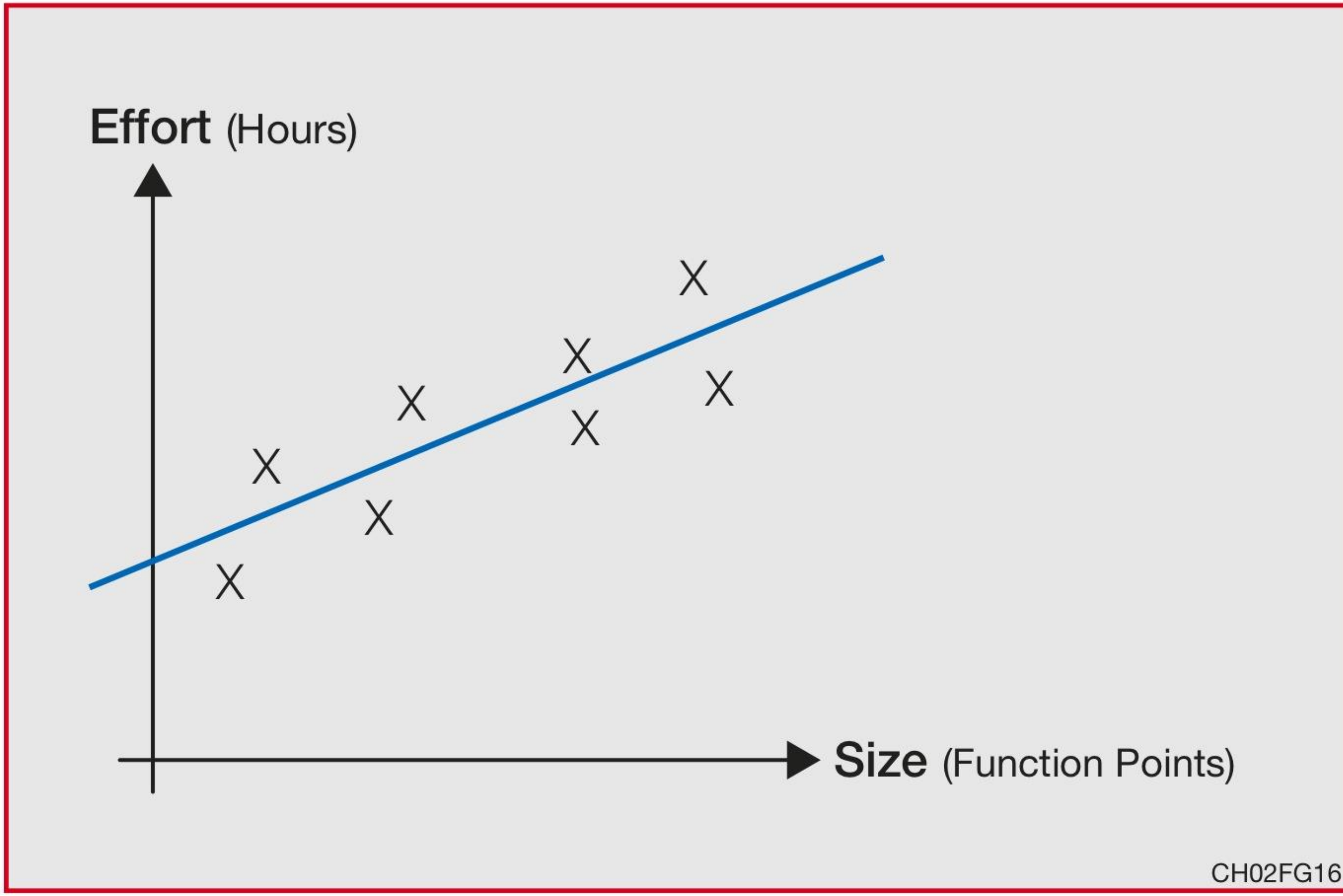


Figure 2.16 An homogeneous size-effort dataset in software engineering.

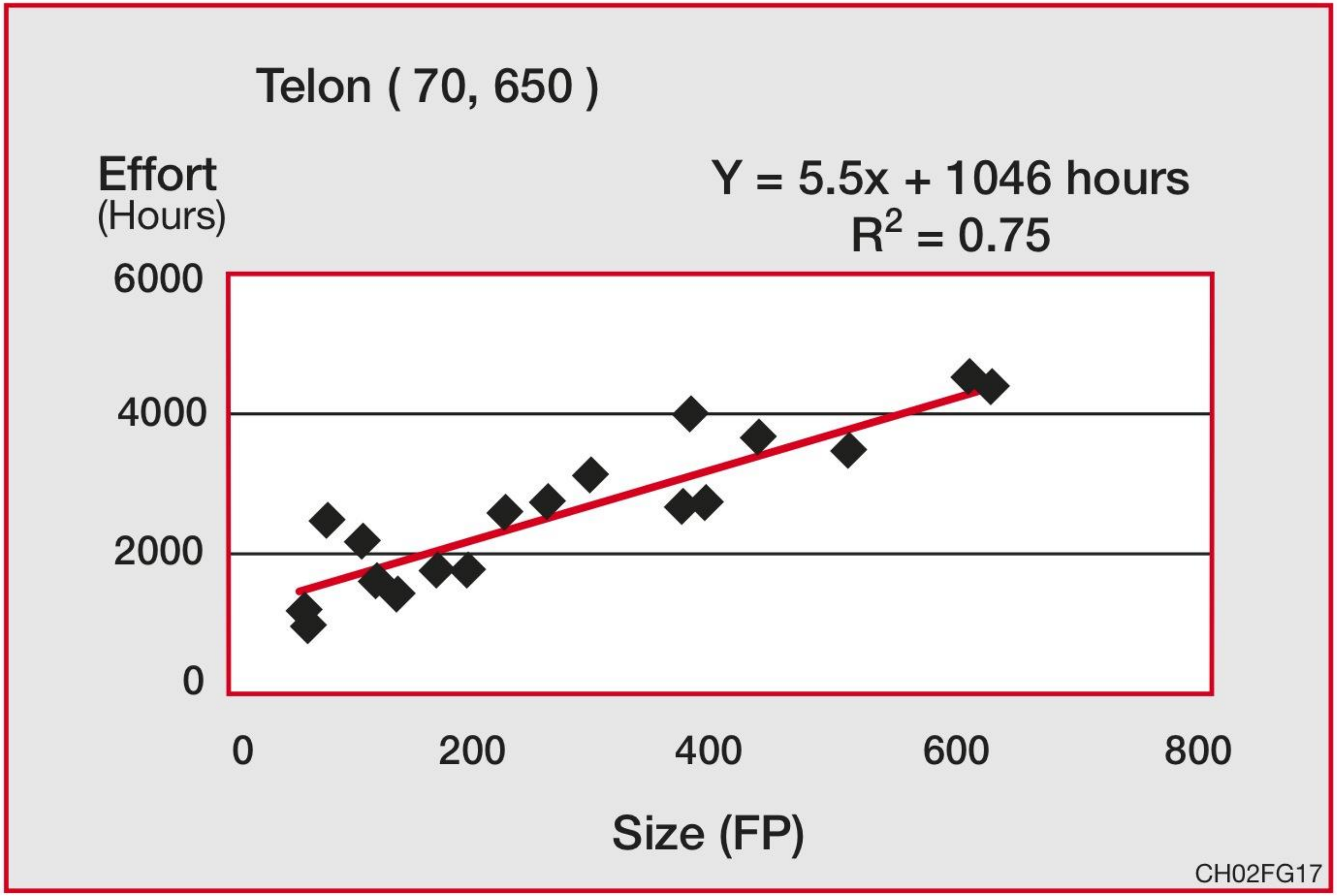


Figure 2.17 The TELON dataset in the ISBSG 1999 Release
(Abran, Ndiaye, Bourque, 2007)

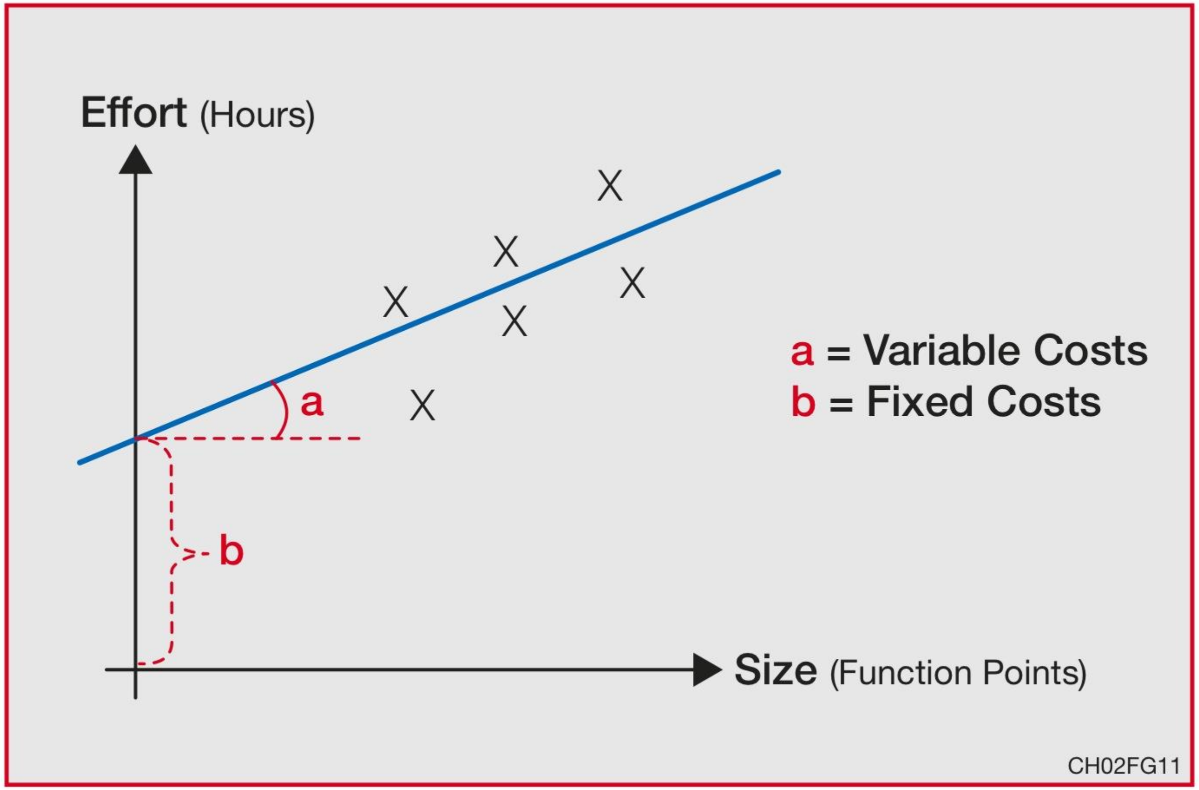


Figure 2.11 Model with a fixed and variale costs.

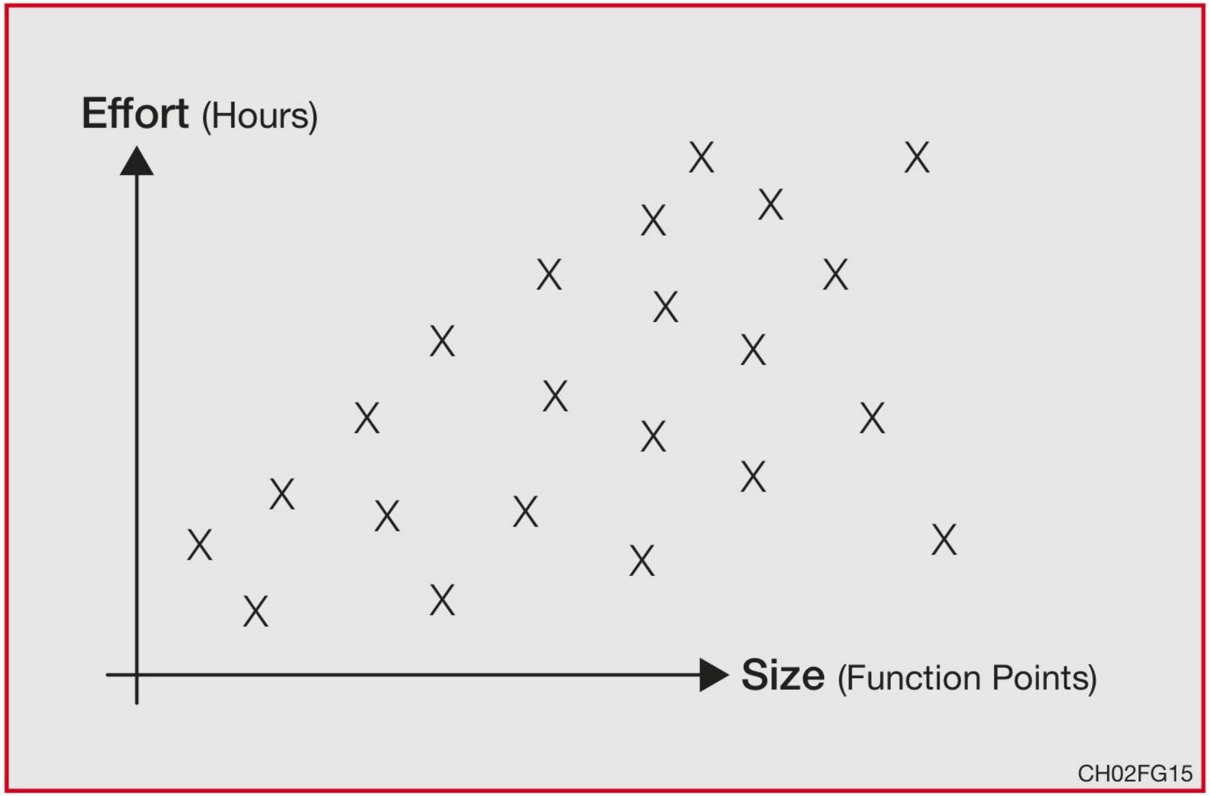
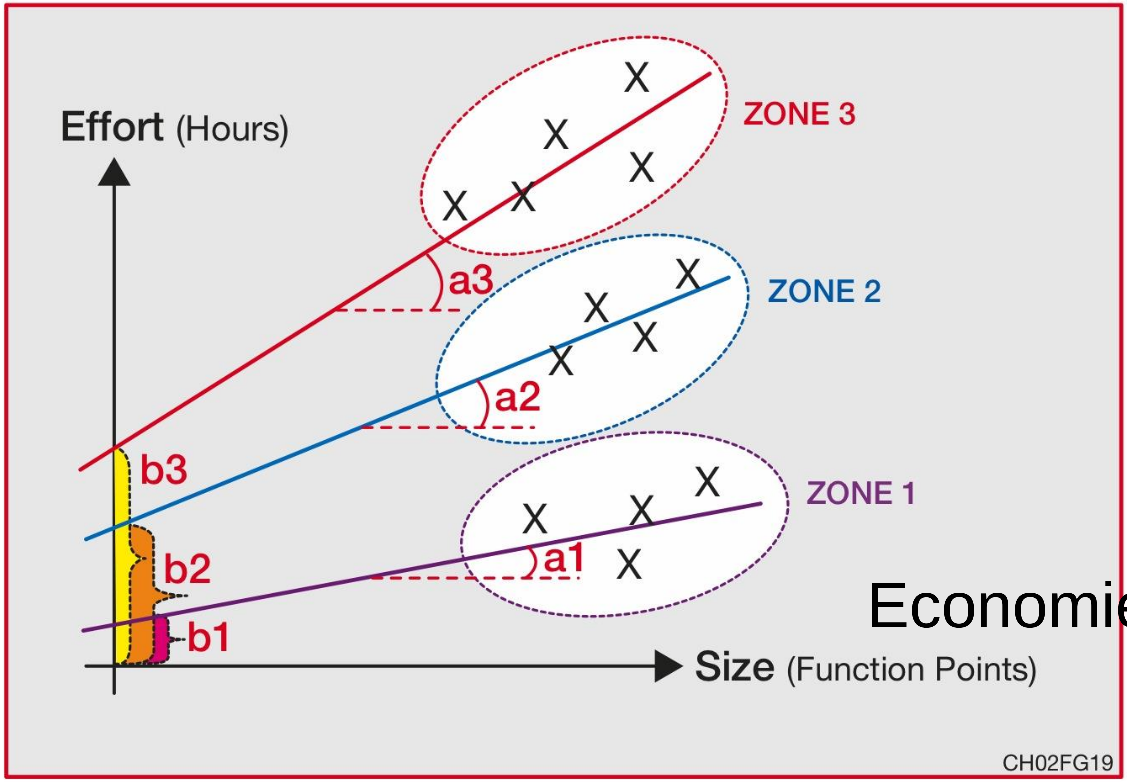


Figure 2.15 Wedge-shaped dataset in software engineering.

Diseconomies of scale



Economies of scale

Figure 2.19 Wedge shape with 3 data subsets with economies/diseconomies of scale.

- Projects from a financial governmental organization

Projects from a financial governmental organization

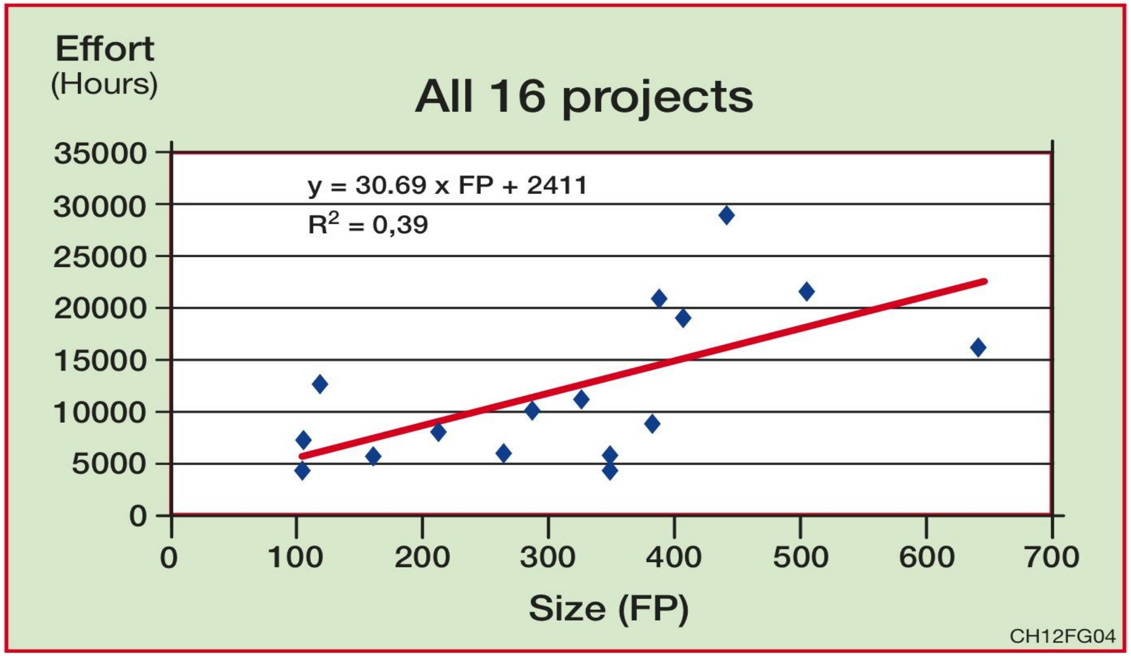


Figure 12.4 The organization's production model.

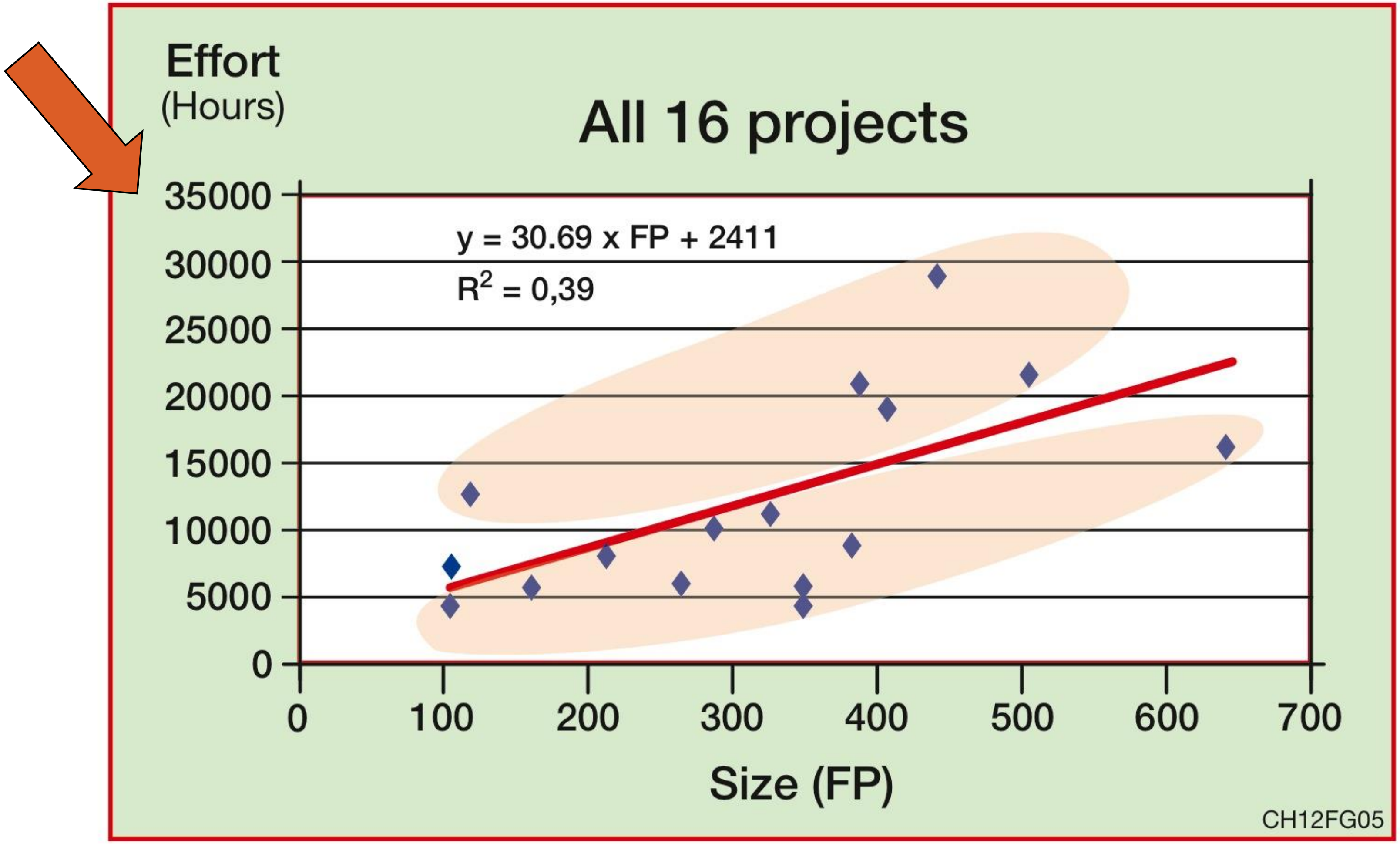


Figure 12.5 The two subsets of projects within the single dataset.

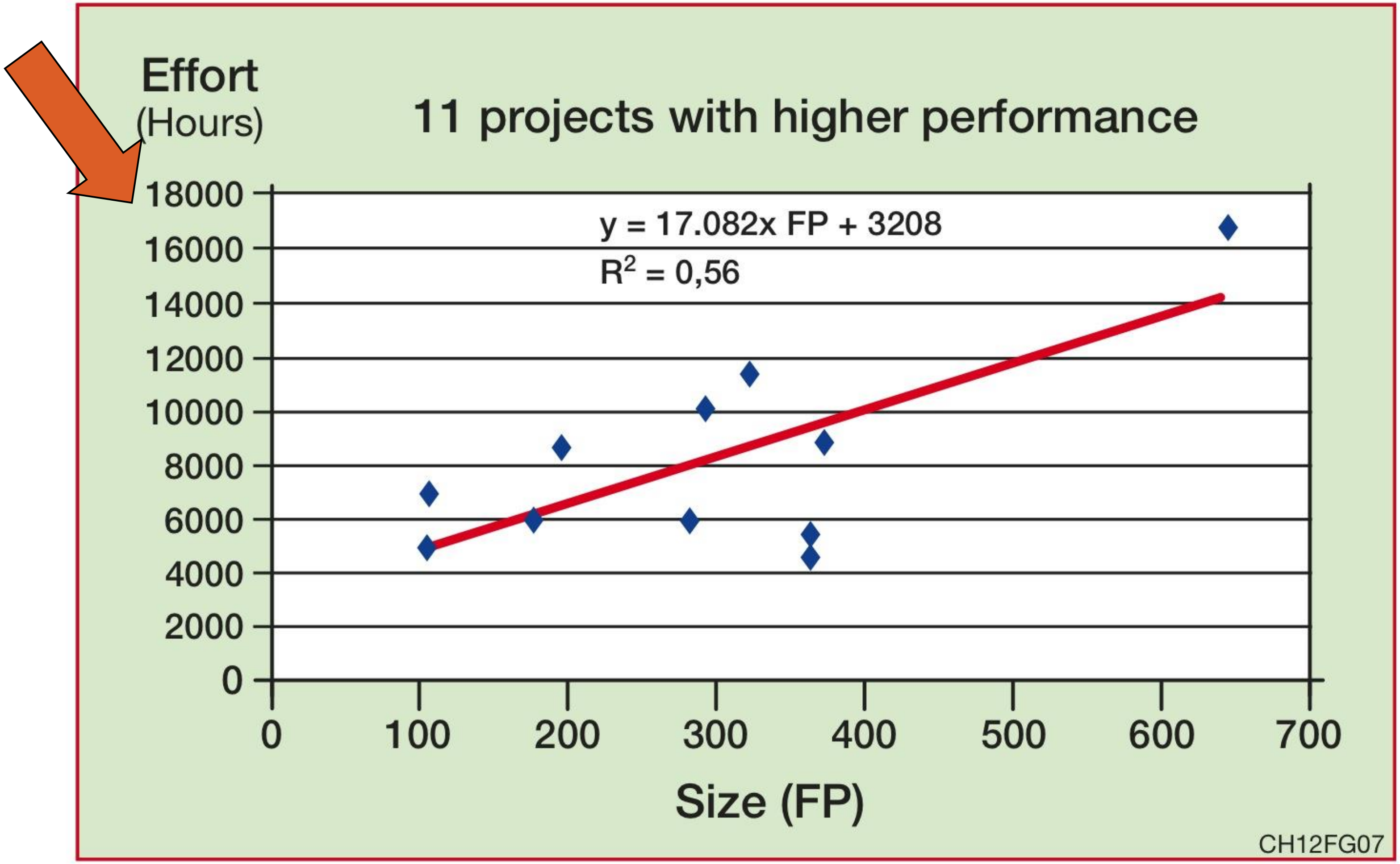


Figure 12.7 Most productive projects.

Causes: Schedule compression, users changing their minds, integrated applications…

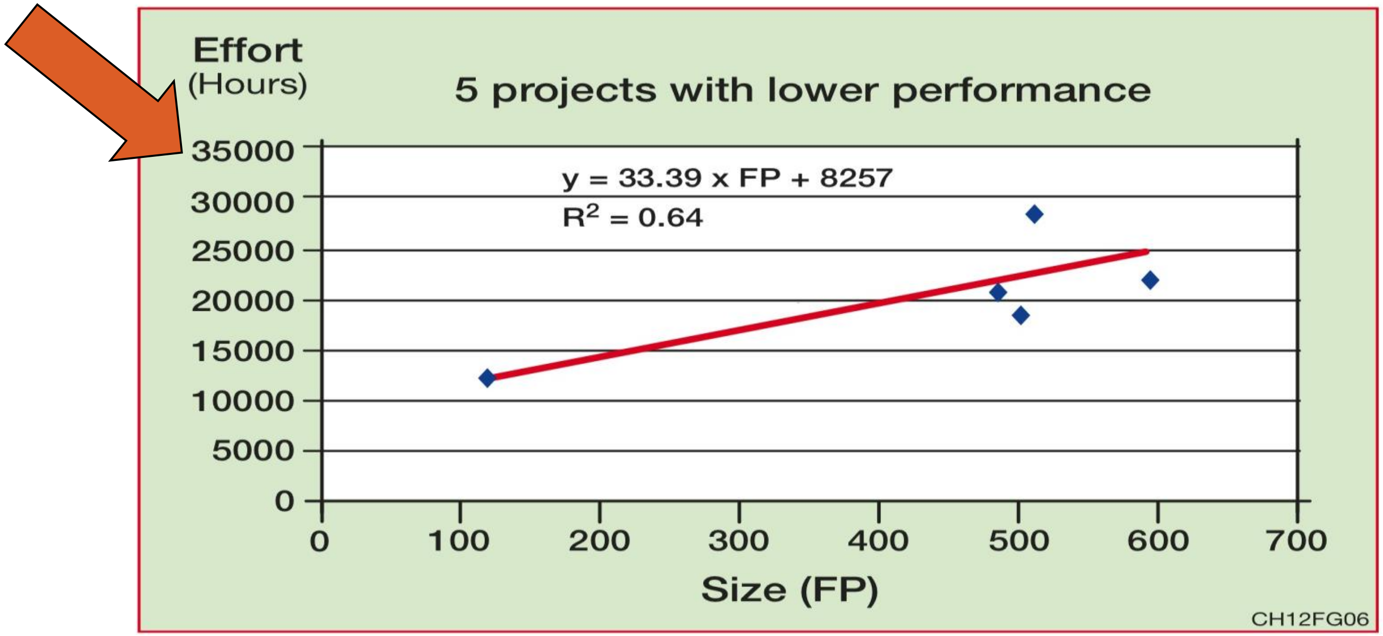


Figure 12.6 Least productive projects.

Which estimation model to use in which contexts?

A Management Decision!

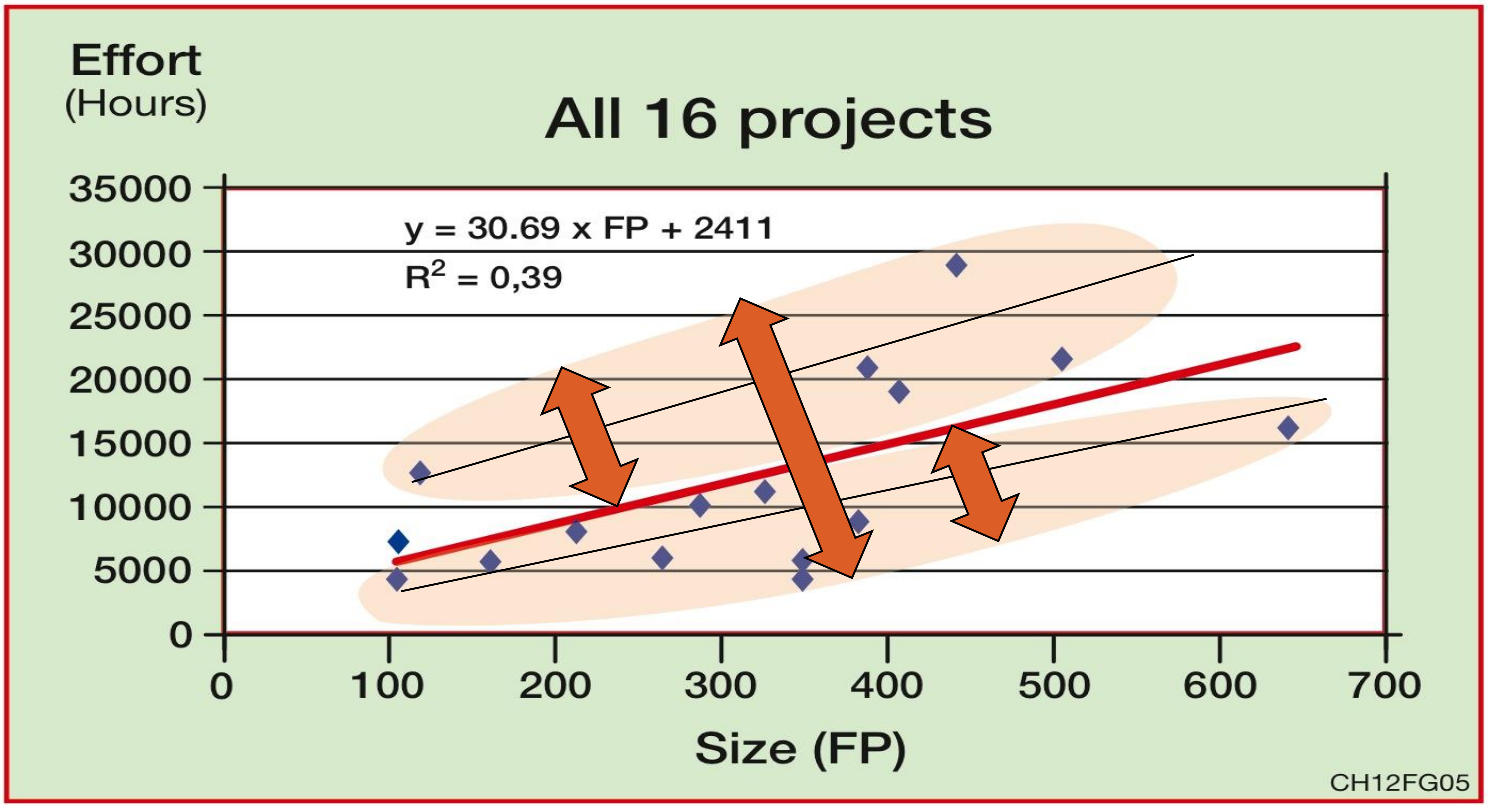


Figure 12.5 The two subsets of projects within the single dataset.

Estimation Models based on economic concepts

A large scale success story:

- Embedded software domain
- Requirements & Specifications:
  - In-house
  - Model driven using Matlab-Simulink
- Software development:
  - Outsourced across the world
  - with qualified suppliers

Estimation Models based on economic concepts

Initial productivity models developed with 20 to 30 projects **<u>for each software supplier</u>**:

- Based on 2nd generation COSMIC size method
- R^2 within the 0.8 to 0.9+ range
- MMRE varies for each supplier
- Info on both fixed & variable costs used to compare suppliers:
  - Simple models that 'talk' to managers based on international standards – No 'black boxes' & game playing with numbers!
- Info on variance to negotiate next projects

Automated COSMIC measurement

- + 300 projects to size and estimate each each at rush time every yeat
- Investment in automation of functional size measurement (with a PhD student)
  - Automation results verified with duplicate measurements over +70 projects (manual & automated).
  - Accuracy of size automation:
    - Prototype: 96%
    - Final automation tool: 99+%

Other usages of functional size measurement

- Prediction model of memory size based on the size of the functional specifications

- Balancing the worload within the team of 100 engineers preparing the detailed software specifications for outsourcing

- Setting annual productivity increases to their network of software suppliers
  - as mandated to their hardware suppliers

Lessons learned

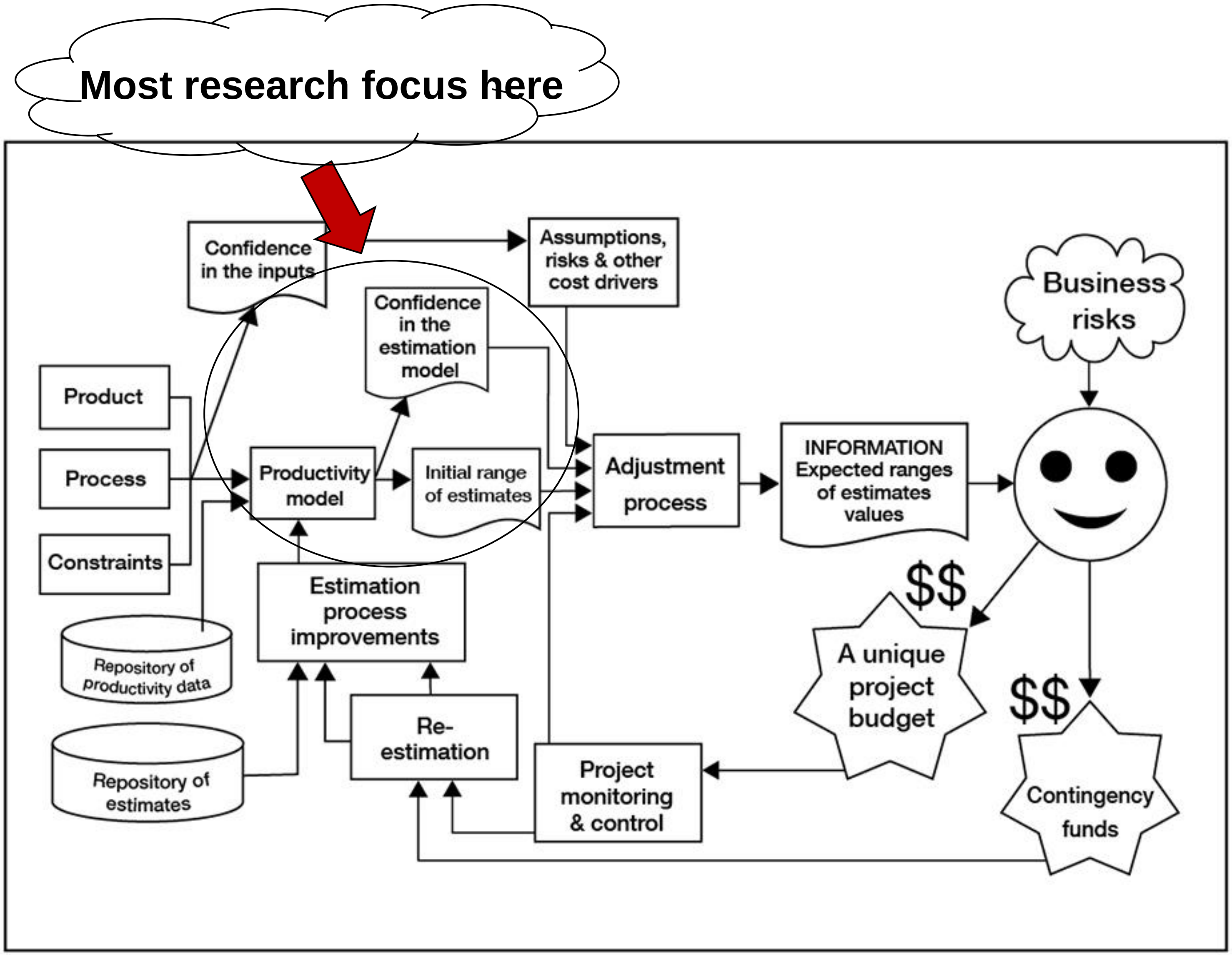
This organization:

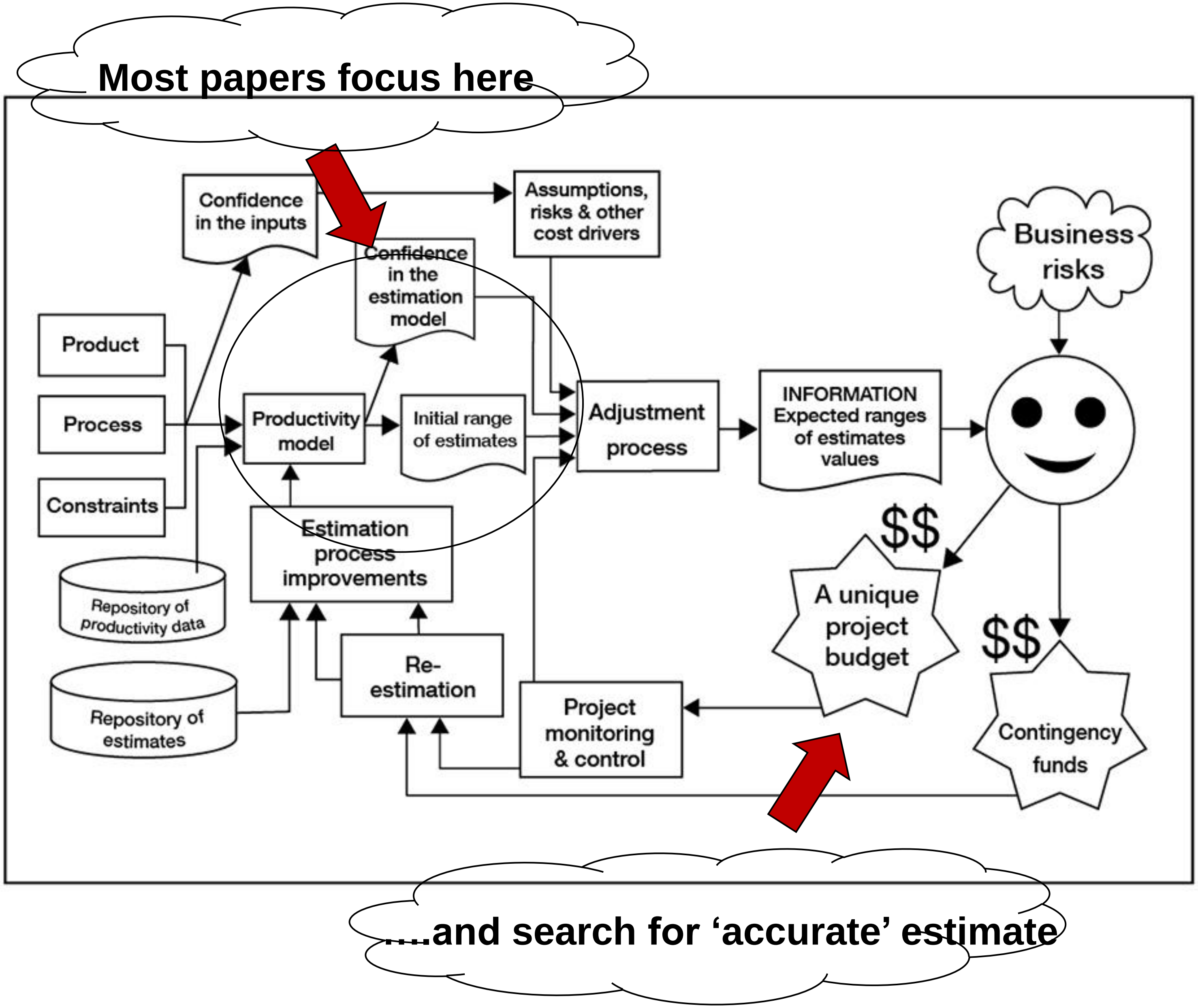
- did not look for miracles (quick & at no cost)!

- They invested time & monies to build a competitive advantage by:
  - Collecting historical data
  - Using standards for measurement
  - Developing minimum statistical skills
  - Being transparent with software suppliers

List of topics

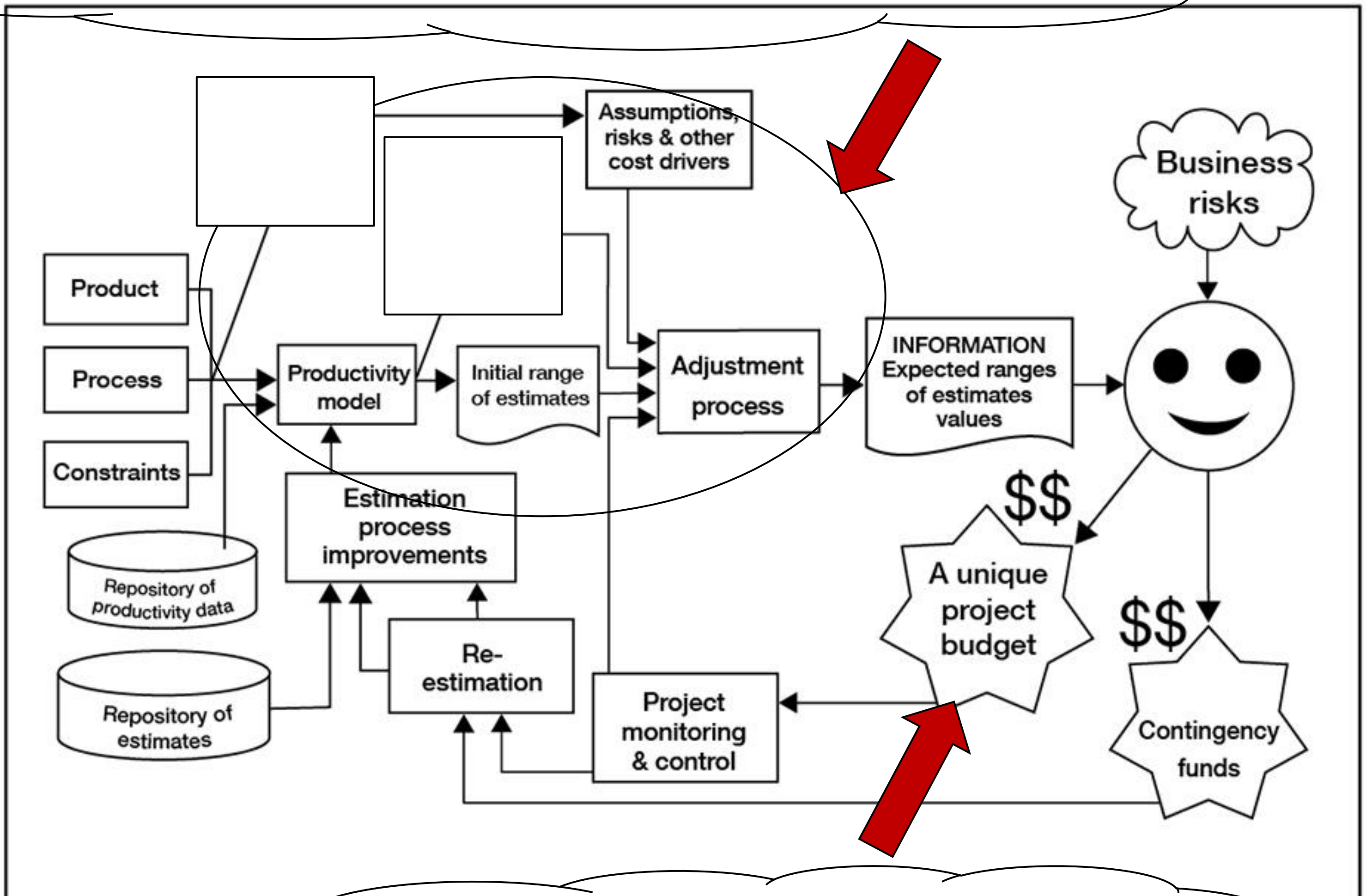
1. Estimation: Craft or Engineering?
2. The estimation phases
3. Economic concepts for estimation models
4. **Orphean research issues**


Most research focus here
Confidence in the inputs
Assumptions, risks & other cost drivers
Business risks
Confidence in the estimation model
Product
Process
Constraints
Productivity model
Initial range of estimates
Adjustment process
INFORMATION Expected ranges of estimates values
Estimation process improvements
Repository of productivity data
Repository of estimates
Re-estimation
Project monitoring & control
$$
A unique project budget
$$
Contingency funds



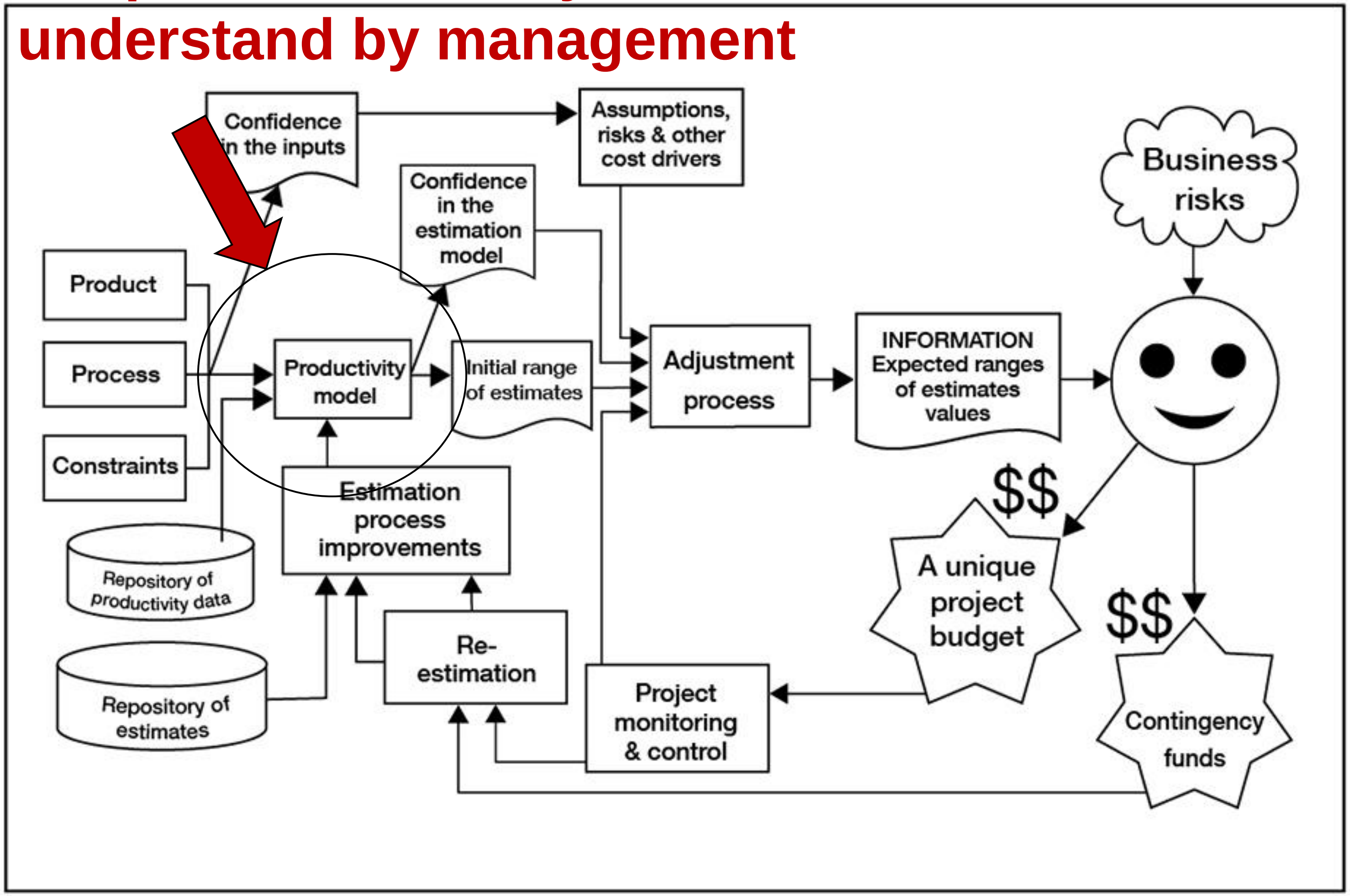
Most papers focus here
Confidence in the inputs
Assumptions, risks & other cost drivers
Business risks
Confidence in the estimation model
Product
Process
Constraints
Productivity model
Initial range of estimates
Adjustment process
INFORMATION Expected ranges of estimates values
Repository of productivity data
Estimation process improvements
$$
A unique project budget
$$
Repository of estimates
Re-estimation
Project monitoring & control
Contingency funds
....and search for 'accurate' estimate

…and many bundle all factors into 'Black boxes'

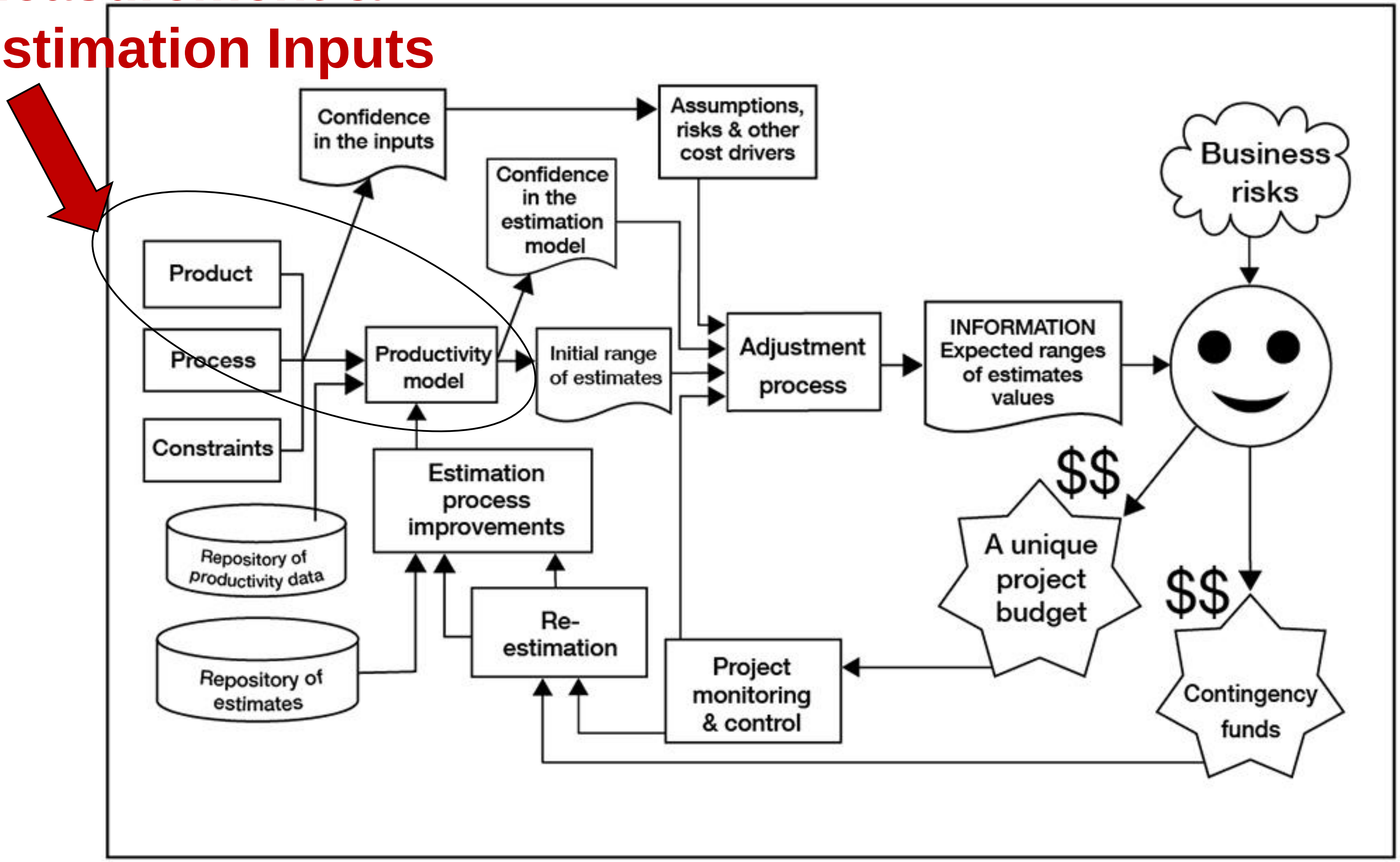


..while searching for 'accurate' estimate

Simple models easy to understand by management



Issues on Software Measurement & Estimation Inputs



Software Size?

- Lines of code

Or

- Function Points:
  - +30 variations
  - & 5 International Standards!

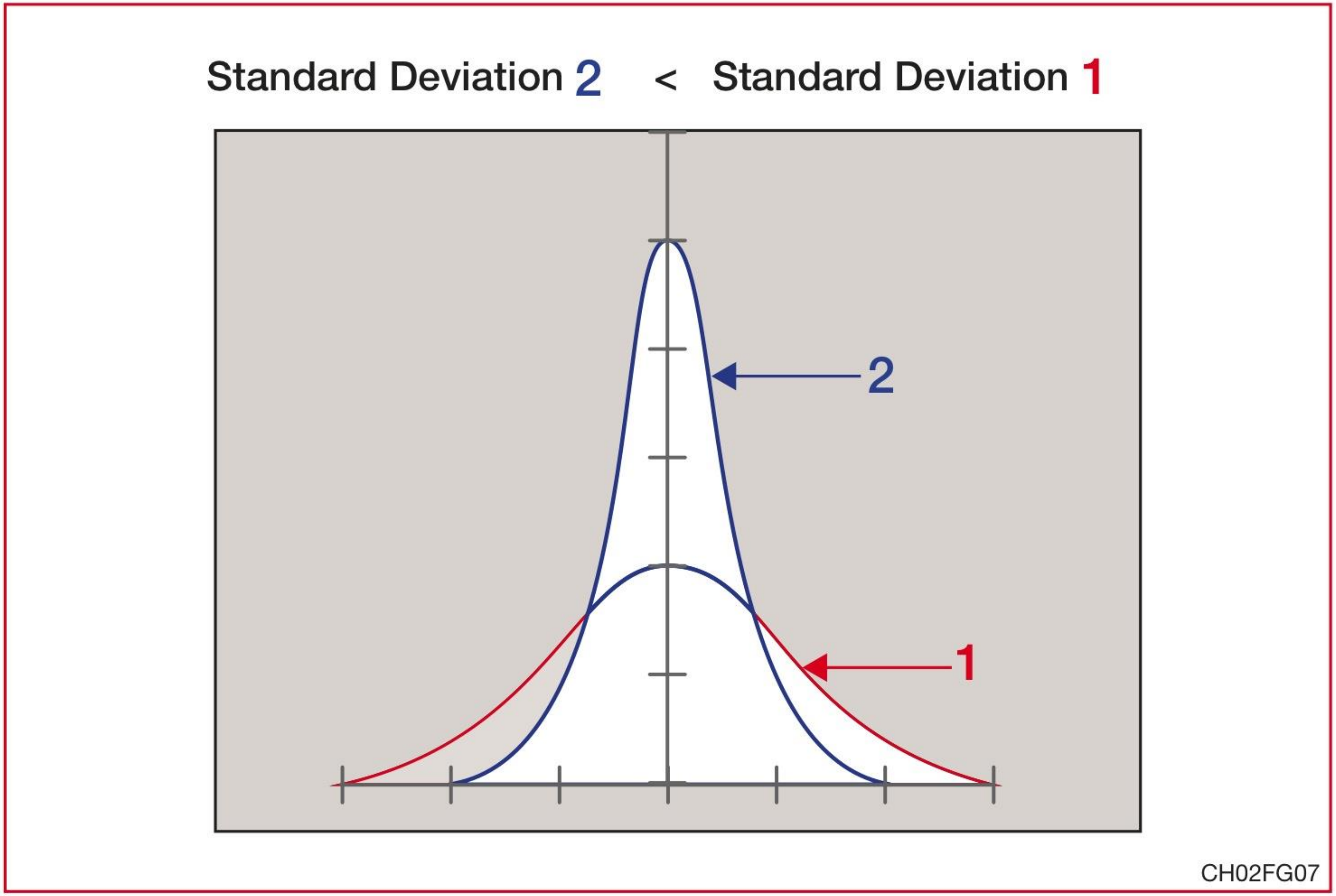


Figure 2.7 The kurtosis in a normal distribution.

Turning dust into gold…

FP to LOC convertion ratios in Estimation Models

- What happened to Ariane 5 spacecraft … and why?

Issues on Estimation Inputs

Which method for software functional size:

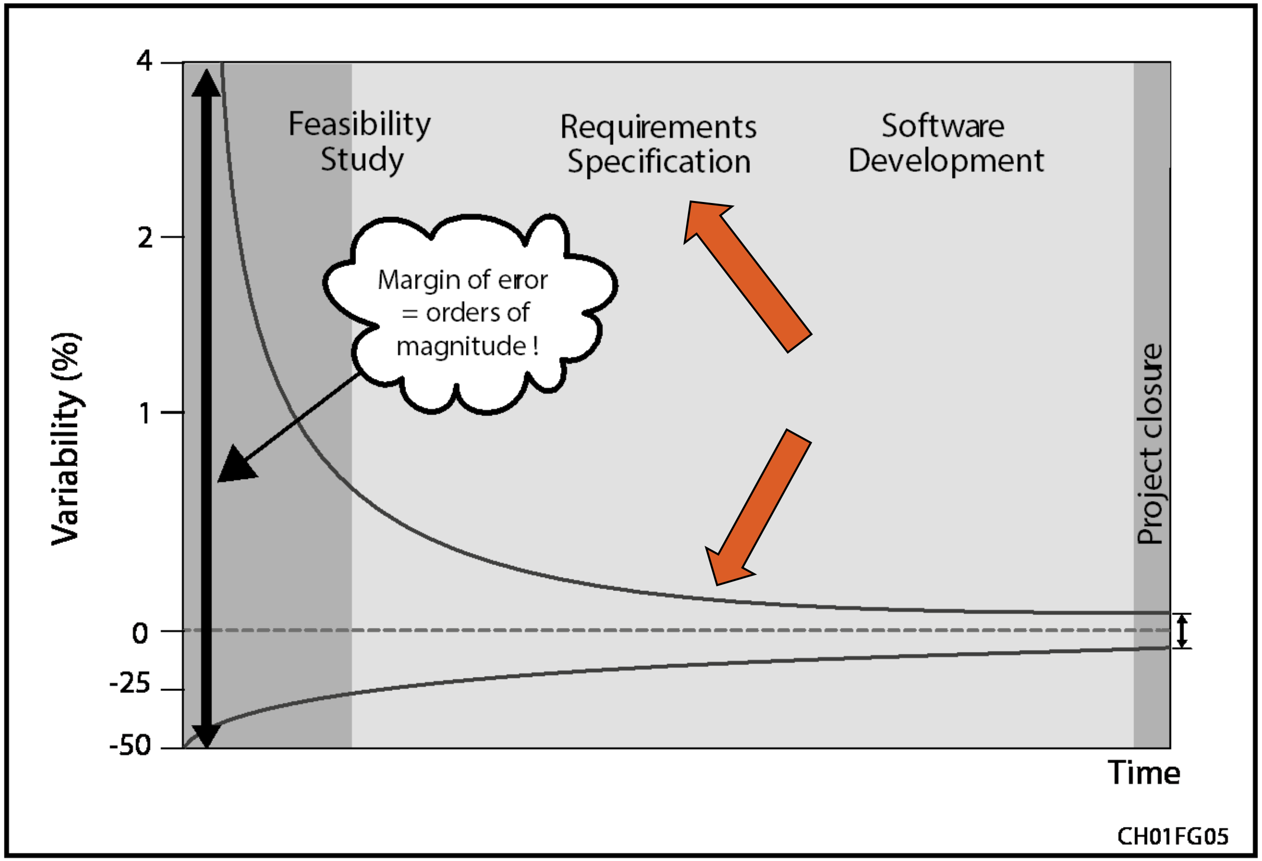
- 1st generation: IFPUG Function Points - 1979
  - Innovator (in 1979, but not in 2004!)
  - Systematic errors! (step function with min & max)
  - Invalid maths!
  - No measurement unit!
  - Still cannot be automated & be compliant after 35 years!
- 2nd generation: COSMIC Function Points - 2003
  - Strenghts based on metrology principles
  - Can be automated & compliant to ISO
  - Applicable across domains
  - Free & + 15 translations

Other Issues on Estimation Inputs

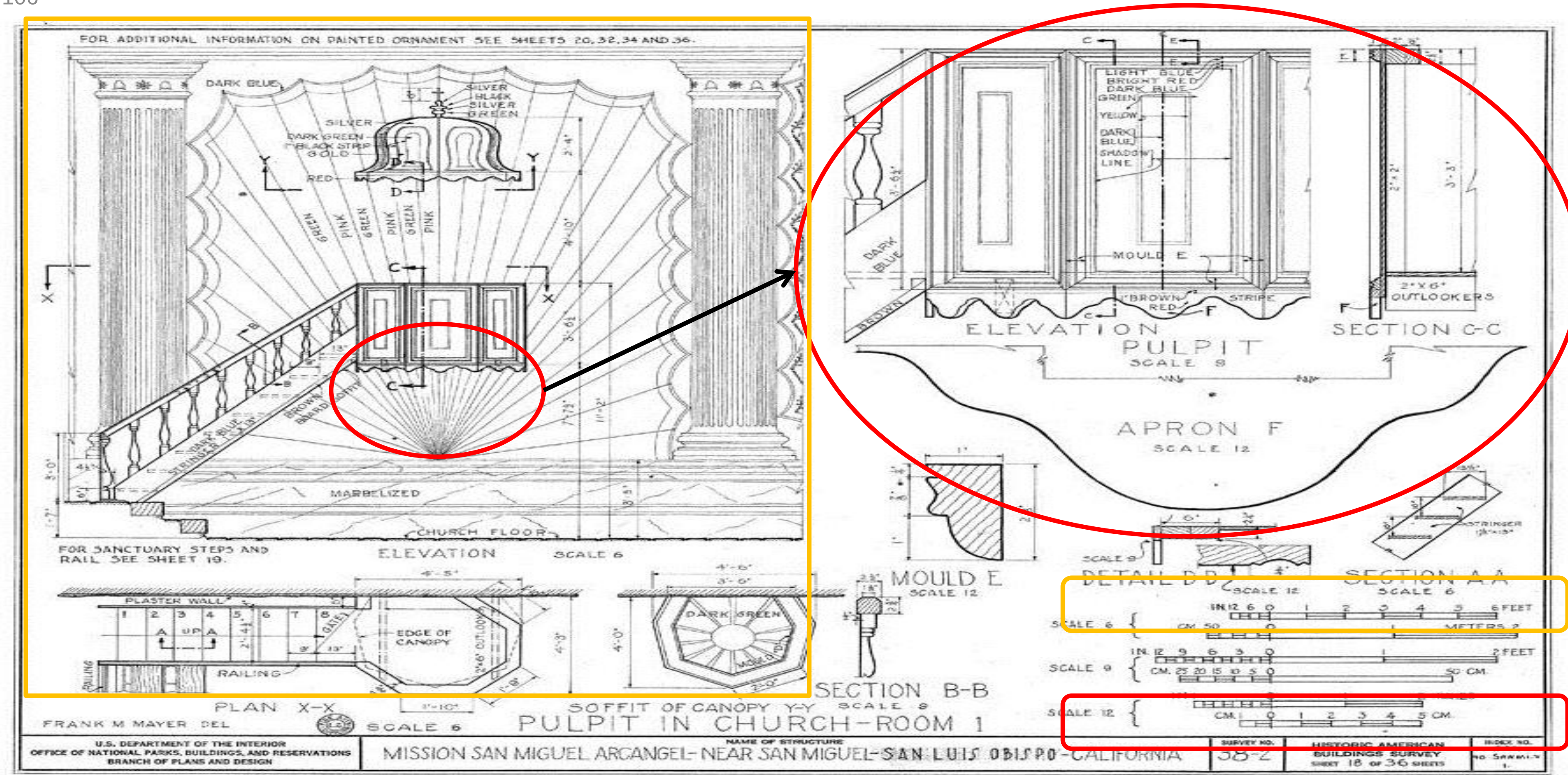
- Unsound sizing methods compounding mistakes:
  - Usecase Points
  - Story Points

- For incomplete software requirements documents, lack of independently verified approximate sizing method

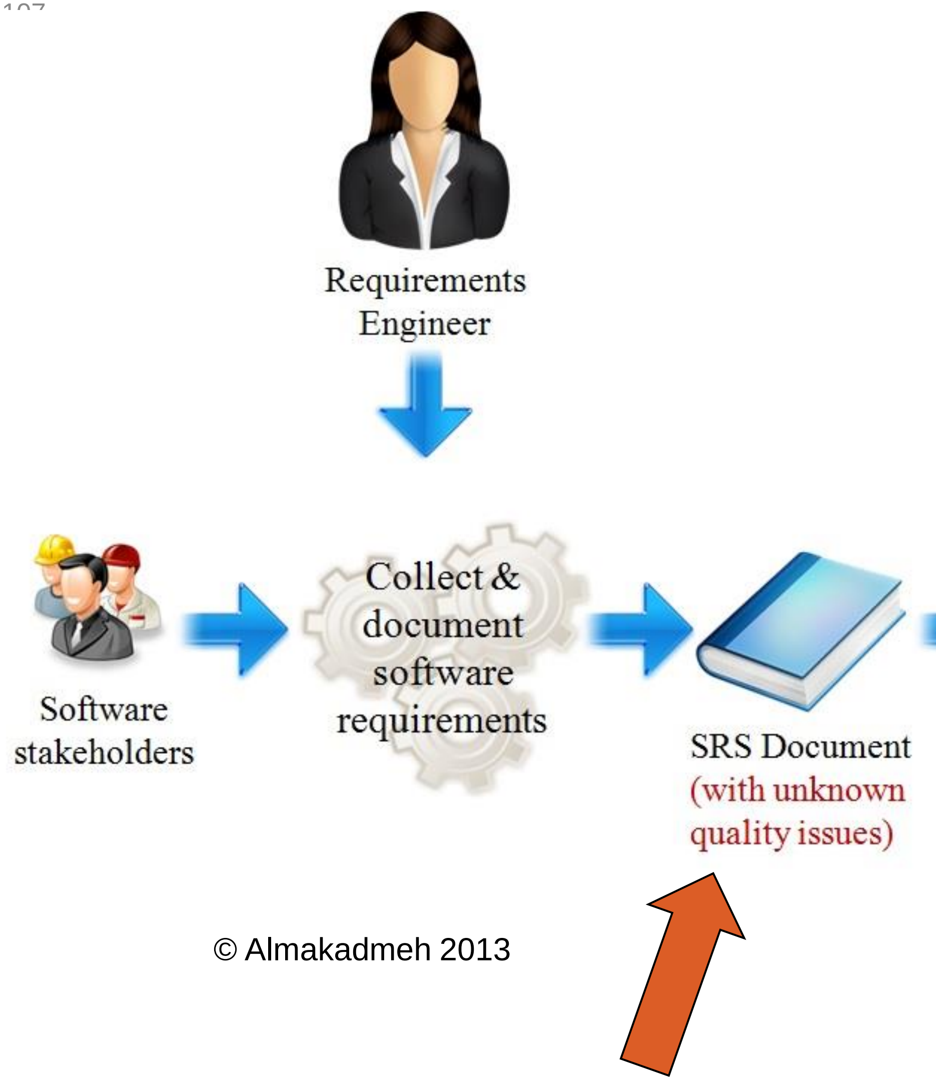
Estimation Models: The Uncertainty Cone: Requirements Specs



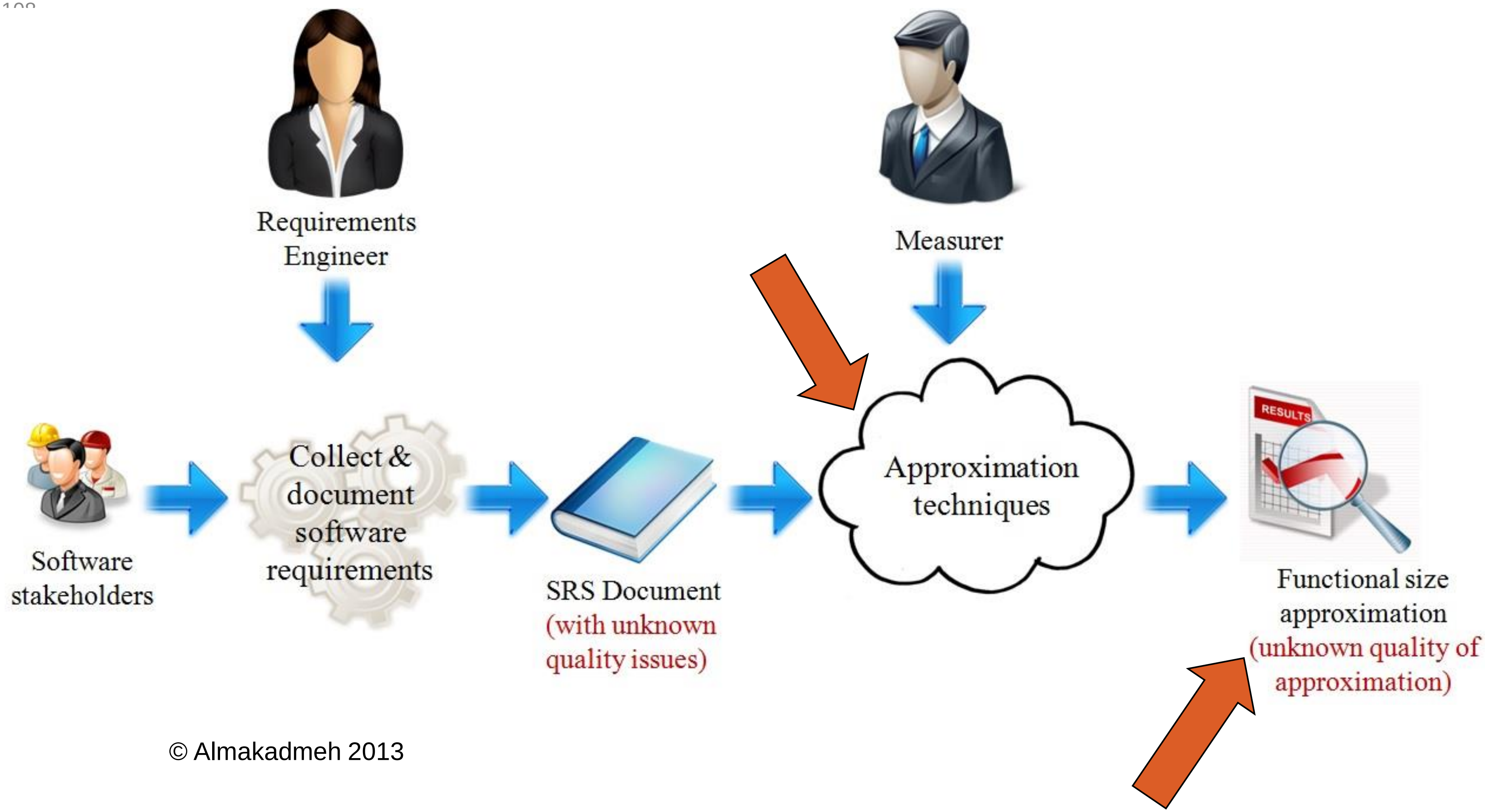
Scales in Plans - Architects & Engineers



Scales in Software Documents?

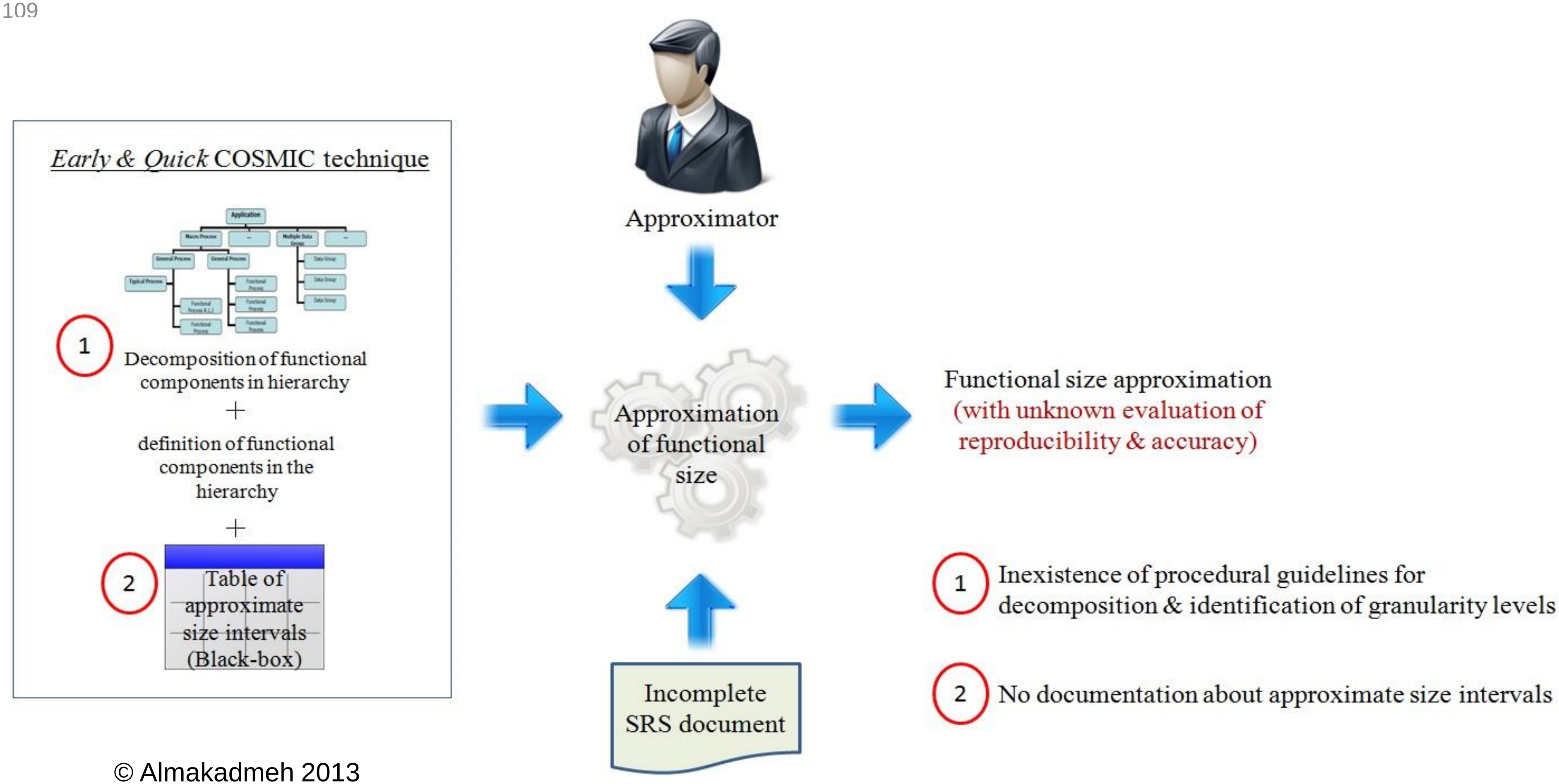


Scales in Software Requirements Texts?



© Almakadmeh 2013

A functional size approximation technique: Unkown Performance..!



© Almakadmeh 2013

An investigation of an existing functional size approximation technique: reproducibility

Difference of functional size approximation

12 Participants

Participant code	Approximate functional size using the *E&Q* COSMIC technique (Min, Most-likely, Max) (in CFP)	Percentage difference in functional size approximation (w.r.t. Most-likely value)
A6	(45, 74, 93)	− 90%
A12	(57, 114, 179)	− 84%
A3	(238, 543, 910)	− 23%
A9	(250, 545, 909)	− 23%
A5	(299, 592, 962)	− 16%
A2	(250, 705, 1250)	0%
A1	(521, 1071, 1616)	+ 52%
A11	(581, 1185, 1972)	+ 68%
A8	(697, 1454, 2472)	+ 106%
A7	(964, 2077, 3450)	+ 195%
A4	(1181, 2369, 3957)	+ 236%
A10	(2265, 4510, 7408)	+ 540%
Minimum		**− 90%**
Maximum		**+ 540%**

An investigation of an existing functional size approximation technique: accuracy

Accuracy of the functional size approximation

12
Participants

© Almakadmeh 2013

Participant code	Approximated functional size using the *E&Q* COSMIC technique in CFP (min, most-likely, max) (1)	Reference functional size for accuracy criteria (2)	MRE calculated using values in (1) and (2) (min, most-likely, max)
A6	(45, 74, 93)	79 CFP	(43%, 7%, 17%)
A12	(57, 114, 179)		(28%, 44%, 126%)
A3	(238, 543, 910)		(200%, 585%, 1047%)
A9	(250, 545, 909)		(215%, 587%, 1046%)
A5	(299, 592, 962)		(277%, 646%, 1113%)
A2	(250, 705, 1250)		(215%, 789%, 1476%)
A1	(521, 1071, 1616)		(557%, 1251%, 1938%)
A11	(581, 1185, 1972)		(633%, 1394%, 2387%)
A8	(697, 1454, 2472)		(779%, 1733%, 3017%)
A7	(964, 2077, 3450)		(1115%, 2519%, 4250%)
A4	(1181, 2369, 3957)		(1389%, 2887%, 4890%)
A10	(2265, 4510, 7408)		(2756%, 5587%, 9241%)
Average MRE on functional size approximations (all 12 participants)			**(684%, 1502%, 2546%)**
Average MRE on functional size approximations (except participants A6 & A12)			**(814%, 1798%, 3041%)**

Estimation Approaches

The '***feel-good***' → dead end!

Quick & Easy…

Building 'good' estimation process & good estimation models

It requires:

- Recognition of uncertainties: how to recognize this and how to deal with it
- The estimator has to provide information, not a single estimate
- The manager has to select a single budget number, and manage risks through contingency planning.
- Discipline, rigor, commitments and $$$

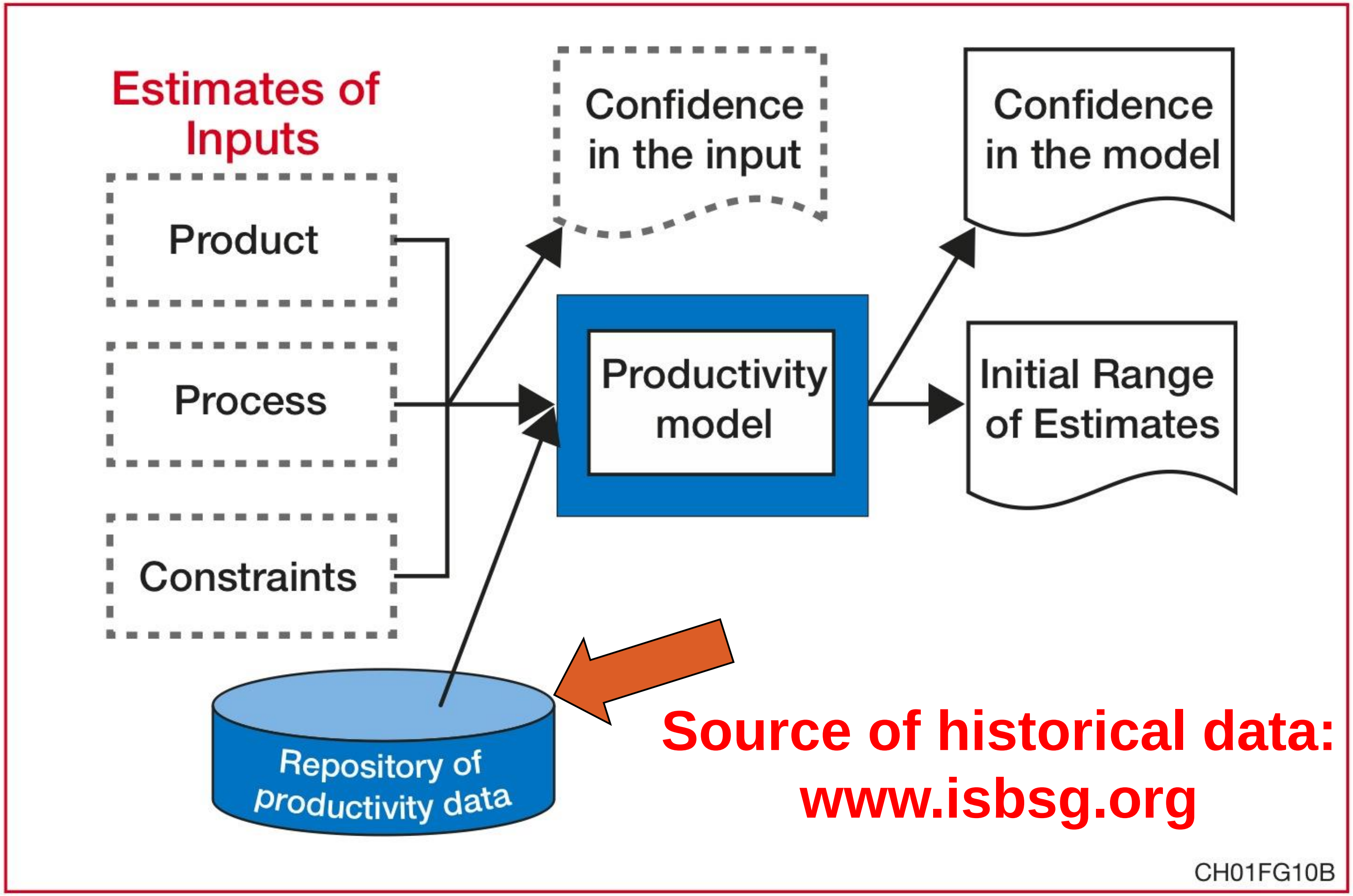


Figure 1.10 Phase B : Execution of the productivity model.

Orphean research issues: Research on software estimation dates back to the early 70's, but much still remain……

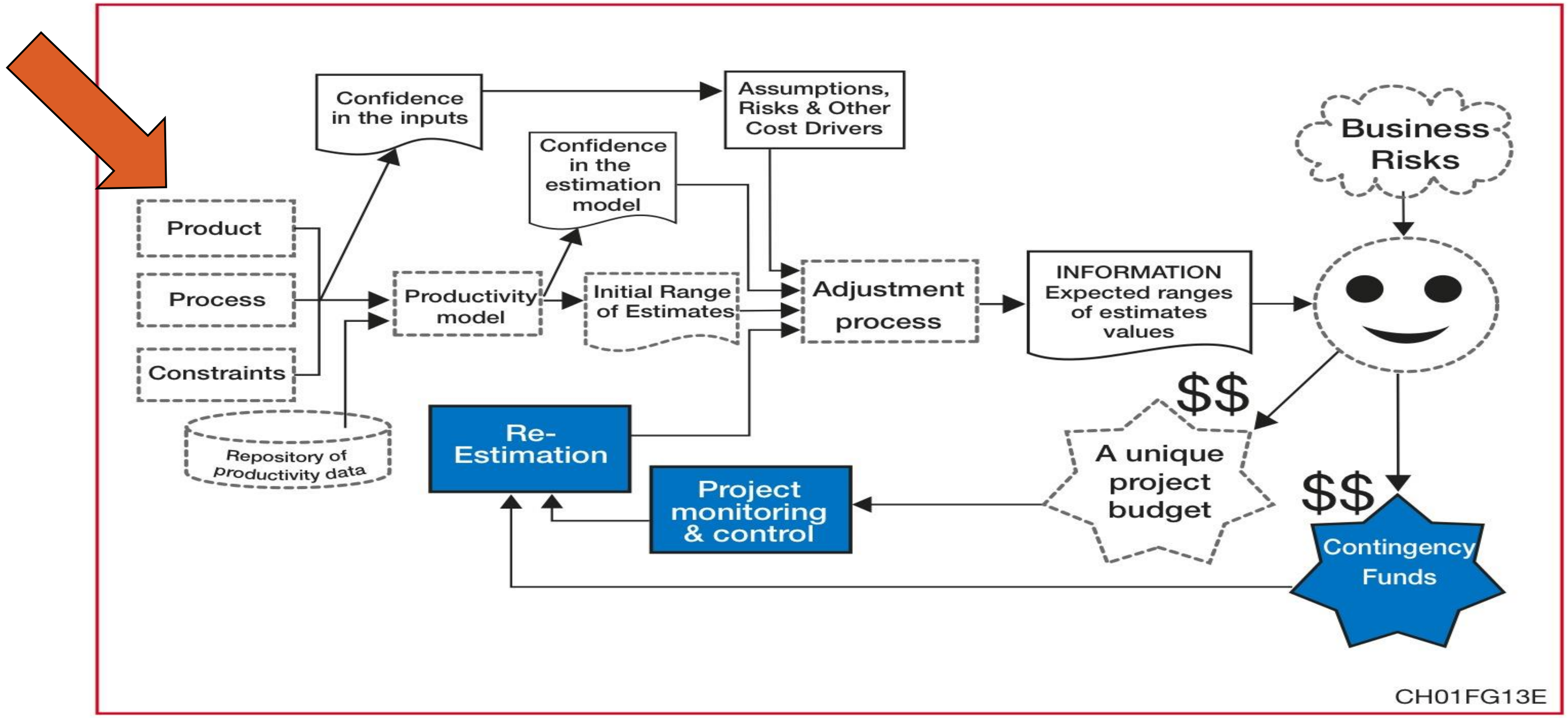


Figure 1.13 Phase E : Re-Estimation.

Orphean research issues:
Research on software estimation dates back to the early 70's, but much still remain……

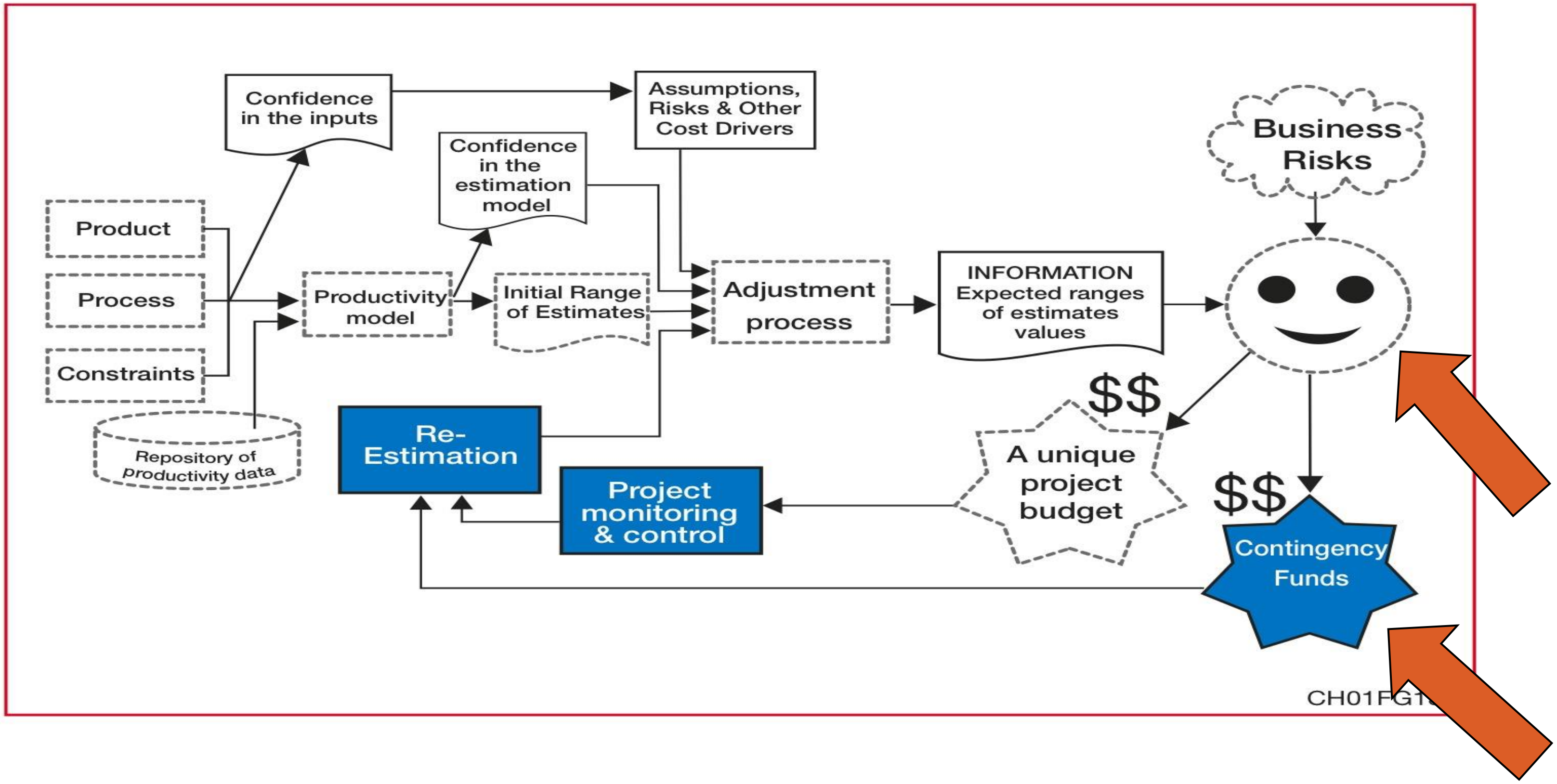


Figure 1.13 Phase E : Re-Estimation.

Orphean research issues:
Research on software estimation dates back to the early 70's, but much still remain……

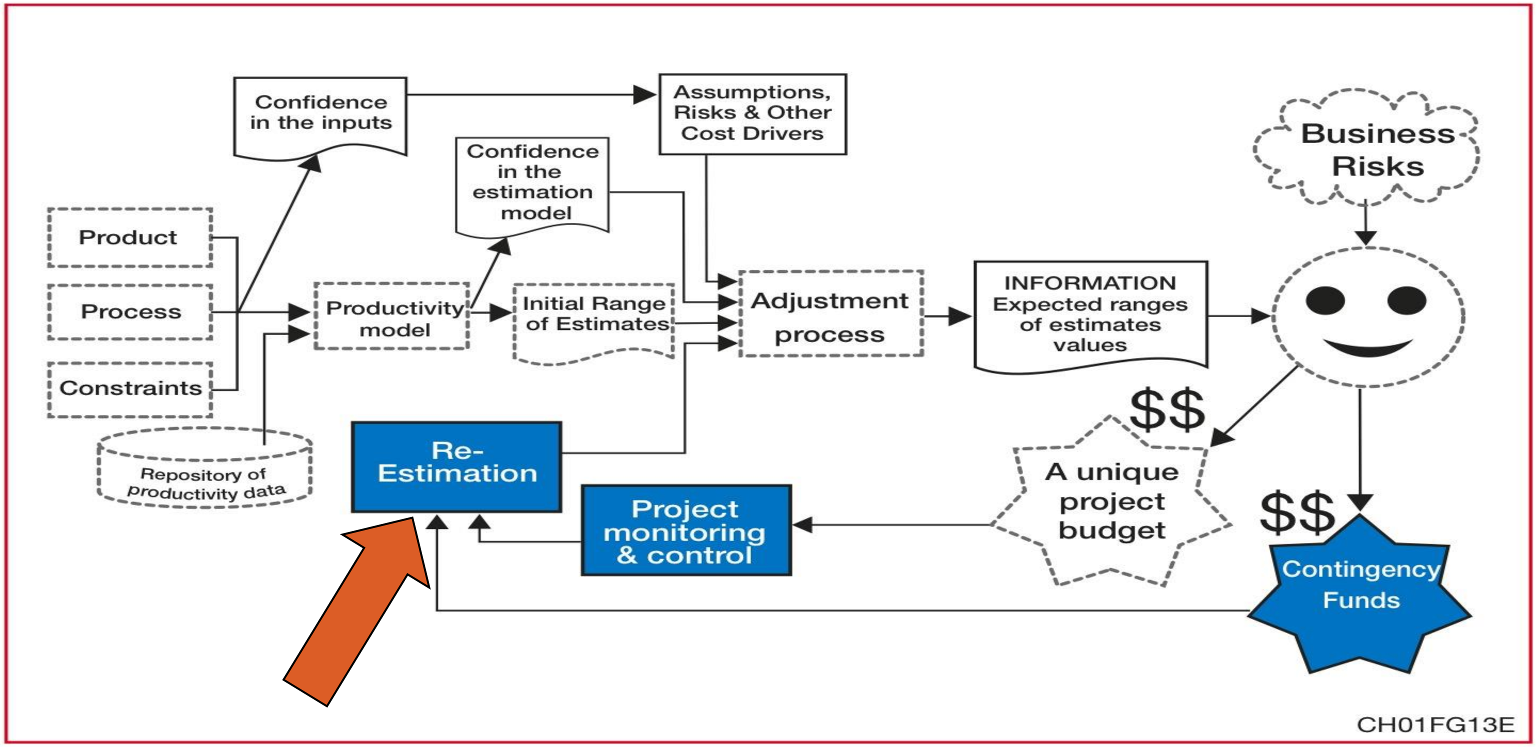
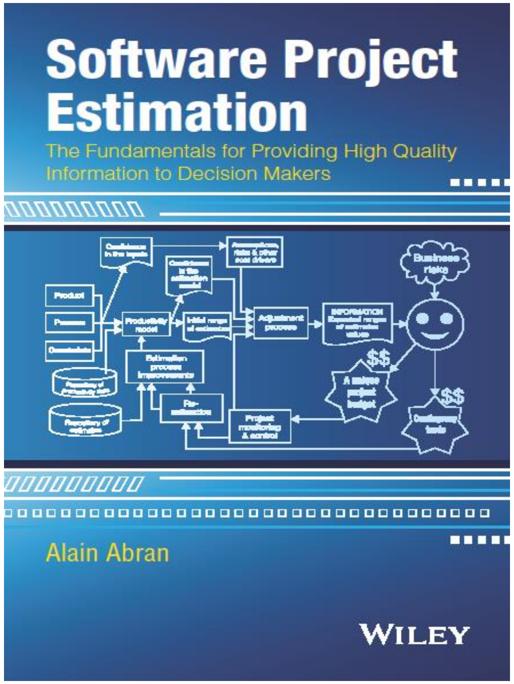
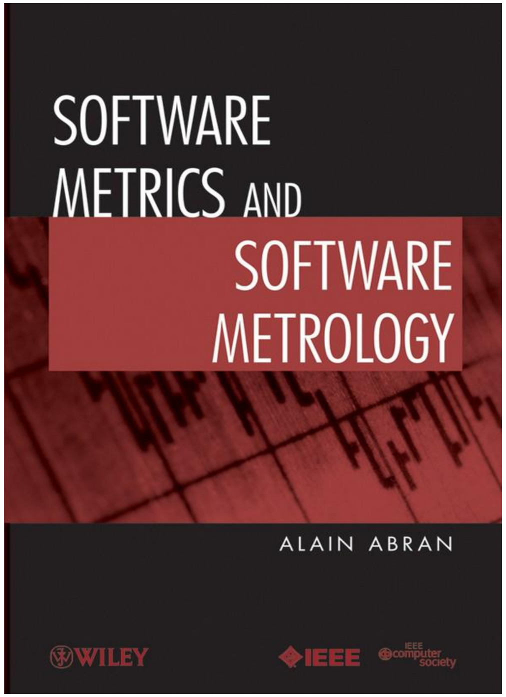


Figure 1.13 Phase E : Re-Estimation.

You want to know more?





alain.abran@etsmtl.ca